Karen Andresa Teixeira Santorum

POR UMA MUDANÇA DE FOCO NA AVALIAÇÃO DA COMPREENSÃO LEITORA

Dissertação apresentada ao Programa de Pós-Graduação em Letras – Mestrado, Área de Concentração em Leitura e Cognição, Universidade de Santa Cruz do Sul, como requisito parcial para obtenção do título de Mestre em Letras.

Orientador: Prof. Drª.Rosângela Gabriel

Santa Cruz do Sul, março de 2007

Karen Andresa Teixeira Santorum

POR UMA MUDANÇA DE FOCO NA AVALIAÇÃO DA COMPREENSÃO LEITORA

Esta Dissertação foi submetida ao Programa de Pós-Graduação em Letras – Mestrado, Área de Concentração em Leitura e Cognição, Universidade de Santa Cruz do Sul – UNISC, como requisito parcial para obtenção do título de Mestre em Letras.

Prof.ª Dra. Rosângela Gabriel

Professora Orientadora

Prof.ª Dra. Márcia Cristina Zimmer

Prof. Dr. José Marcelino Poersch

Certa vez um rei excêntrico em suas fantasias e preocupado com a falta de integração entre seus súditos anunciou que daria a maior e mais bela festa de seu reino.

Todos se apresentaram com seus trajes mais luxuosos. Chegaram cheios de expectativas, pensando nas iguarias que poderiam ser oferecidas naquele ambiente esplendoroso. Entretanto, durante muitas e muitas horas sucedeu-se uma programação cultural de danças, músicas, cantos, jogos refinados e divertidos, sem que nenhum comestível fosse oferecido.

Os convivas foram sentindo uma grande fome que, com o passar das horas, tornou-se incontrolável e intolerável, suscitando um grande sentimento de contrariedade em todos eles. Só então o rei convidou-os para uma estranha e surpreendente refeição. Apenas uma deliciosa e rica sopa servida num grande caldeirão, com colheres de metal de mais de um metro de cabo e apenas uma pequena guarnição de madeira na extremidade para que se pudesse segurá-la. Não havia pratos. Em sucessivas e desesperadas tentativas, os convivas tentavam levar a sopa à boca. Era impossível. O desespero já dominava o ambiente quando um dos convidados teve a idéia de levar sua colher à boca de seu vizinho. Estava desvendada a chave. Todos passaram a servir-se mutuamente, até saciarem a fome. E compreenderam então a intenção do rei. Certos empreendimentos têm resultado muito melhor com o concurso da cooperação mútua, do trabalho em equipe.

(História da tradição Sufi)

À minha mãe, parceira incondicional de minha jornada cognitiva, que viabilizou a realização de mais esse grande empreendimento; ao meu pai (*in memorian*) que sempre respeitou e apoiou meus ideais; à minha irmã, que além de me mostrar o caminho, auxiliou na caminhada. A eles dedico minha dissertação.

AGRADECIMENTOS

Certamente a conclusão deste empreendimento não teria sido a mesma se não fosse pela cooperação recebida e pelo trabalho em equipe aqui realizado. Sinto-me imensamente feliz ao chegar neste momento com a certeza de que seria necessário, pelo menos um capítulo, para poder nomear, uma por uma, as pessoas que contribuíram trazendo suas bagagens para esta caminhada. Mas quero agradecer explicitamente algumas pessoas que participaram ativamente na construção de sentidos de meu texto-sonho.

À minha orientadora, Profª. Drª. Rosângela Gabriel, por compartilhar comigo seus conhecimentos prévios e, engramada nesses, a paixão pela lingüística. Por sua capacidade de encontrar sentido em minhas idéias. Por ter me revelado em detalhes o universo da pesquisa científica, do qual eu não quero mais sair. Pela experiência maravilhosa obtida no estágio em psicolingüística. Por ter compartilhado também sua família, seu carinho, pelo encorajamento, enfim, pela parceria, que espero não termine por aqui.

Ao Programa de Pós-graduação em Letras da Universidade de Santa Cruz do Sul pela oportunidade de tantas aprendizagens e pela concretização desse ideal. À Lucilene, secretária do mestrado, por sua paciência e auxílio durante a caminhada.

À CAPES pela concessão da bolsa, condição *sine qua non* para a realização desse *sonho de consumo* que era para mim o mestrado.

À equipe do Comut da biblioteca da UNISC pelas incansáveis buscas por tesouros perdidos.

Ao Departamento de Letras da UNISC pela acolhida e viabilização da proposta de pesquisa liberando o uso do laboratório de informática para a produção de dados.

Aos monitores, Paulo e Rodrigo, do laboratório de informática do departamento de Letras que foram incansáveis auxiliando nas etapas de produção de dados.

À Profª. Ms. Lilian Cristine Scherer (orientadora da especialização que antecedeu este mestrado) por seu incentivo e sugestões tão enriquecedoras.

Ao amigo e inspirador Prof. Ms. Edgar Pacheco por suas contribuições tão significativas em meu percurso.

Aos meus amigos (ex) alunos, e aos demais leitores, que consentiram em parar com suas atividades a fim de serem os produtores dos dados desta pesquisa, para os quais e
é um prazer e com os quais aprender é um contínuo.

Ao prof. Renato Michel, que me despertou a paixão insuspeitada pela estatística, cuja paciência foi *estatisticamente significativa*.

Ao amigo Charles Müller por sua imensa generosidade e acolhida da proposta, viabilizando a mesma ao tornar a OUR TOOL uma realidade. O envolvimento, dedicação e carinho na criação da *nossa ferramenta* foram decisivos.

Aos colegas da *tribo*, em especial às lindas *super* parcerias formadas; à comunidade virtual criada, que agia terapeuticamente, despertando uma criatividade inusitada nas altas horas de trabalho na madrugada.

Ao Mozart e colaboradores pelos momentos de liberação dos *espaços mentais* oportunizados por suas sinfonias.

Por fim, agradeço de modo todo especial à minha mãe, Lurdes, por seu apoio integral, por comungar dos meus sonhos e pelas várias panelas de brigadeiro que me servia e que foram responsáveis por manter minha atividade neuronial quando tudo parecia estagnado. Por seu incentivo e exemplo de garra. Ao meu pai, José Alberto, que continua presente em minha vida, iluminando caminhos. À minha irmã, Katia, por seu estímulo constante e por ter viabilizado uma navegação sem limites pelos mares da *internet* em busca de um precioso referencial teórico. Pelos momentos de reflexão no novo espaço em comum da psicolingüística. A Deus por ter me permitido realizar esse mestrado e por ter me cercado de anjos ao longo da trajetória. Agradeço também a todos esses anjos que direta e indiretamente contribuíram para a realização desse empreendimento, pois, de acordo com a *lente* de Vygotsky (1998, p. 79), “o desenvolvimento da espécie humana e do indivíduo dessa espécie está, pois, baseado no aprendizado que sempre envolve a interferência, direta ou indireta, de outros indivíduos e a reconstrução pessoal da experiência e dos significados”.

Obrigada a vocês pelas transformações em mim!!!!!!!!!!

Assim como o desenvolvimento da linguagem oral fez de nós uma espécie única, a aquisição da linguagem escrita modificou nossa espécie de inúmeras formas. Ela modificou nossa capacidade cognitiva (Olson, 1980); ela modificou o tipo de conhecimento que nós poderíamos ter (Havelock, 1976); e ela pode até mesmo ter mudado nossa neuroanatomia (Geschwind, 1974). Talvez nenhuma outra atividade humana tenha alterado o curso do desenvolvimento individual e cultural (Wolf et al., 1998) (AEBERSOLD e FIELD, 1997) [1].

[1] No original: "Just as the development of oral language made us a unique species, the acquisition of written language changed our species in numerous ways. It changed our cognitive capacity (Olson, 1980); it changed the kind of knowledge we could have (Havelock, 1976); and it may even have changed our neuroanatomy (Geschwind, 1974). Perhaps no other human activity has altered the course of individual and cultural development (Wolf et al., 1998)" (AEBERSOLD e FIELD, 1997)

RESUMO

O estudo da leitura enquanto processo de construção de sentidos se apresenta como um caminho para prosseguir nas investigações acerca da cognição. Os fatores envolvidos no processo da leitura são inúmeros, uma vez que a compreensão, como processo mental, se constitui num fenômeno complexo envolvendo uma constelação de capacidades e habilidades cognitivas inter-relacionadas, variando desde mapeamentos mecânicos a manipulações mais conceitualmente sofisticadas como raciocínio e inferência. O estudo dos fatores envolvidos na compreensão assume o importante papel de permitir o planejamento de intervenções mais eficazes, destinadas não apenas a desenvolver a compreensão, mas também a avaliá-la. A ação avaliativa, quando feita a partir de um referencial teórico produtivo, pode informar muito a respeito do processo mental e da autêntica habilidade de compreensão do candidato. A ação avaliativa abrange justamente a compreensão do processo de cognição. Ao discutir aspectos subjacentes à avaliação da compreensão leitora, especificamente em língua estrangeira, nesse caso língua inglesa, a presente pesquisa buscou investigar se o nível ou níveis de compreensão textual alcançados pelos sujeitos ao lerem um texto eram suficientes para executarem as tarefas do teste. Os sujeitos foram divididos em dois grupos e submetidos a dois tipos de avaliações com abordagens diferentes: uma padronizada do tipo múltipla-escolha, nos moldes do teste TOEFL (*Test of English as a Foreign Language*) e outra alternativa, especialmente desenvolvida para a pesquisa, nos moldes da taxonomia de Barret (1968). Após cada avaliação foi pedido um reconto do texto lido e foi realizado o monitoramento das estratégias de leitura utilizadas. Os resultados reportam que, no caso do TOEFL, entre os dados da amostra, não houve uma correlação entre os níveis de compreensão textual alcançados pelos sujeitos e o resultado no teste. Em outras palavras, foi constatado que através de um comportamento específico de leitura, que faz uso de leitura estratégica, é possível responder satisfatoriamente às questões do teste, atingindo um escore alto, sem necessariamente ter atingido um nível de compreensão que garanta uma representação em profundidade do texto. Tal fato ressalta a importância de uma reflexão sobre as decisões a serem tomadas com base nos resultados dos testes.

Palavras-chave: cognição, compreensão leitora, avaliação

ABSTRACT

The study of reading as a process of building meaning appears as a way to continue the investigations concerning cognition. The factors involved in the reading process are countless, once the comprehension, as a mental process, constitutes a complex phenomenon involving a constellation of capacities and interrelated cognitive abilities, varying from mechanical mappings to more conceptually sophisticated manipulations like reasoning and inference. The study of the factors involved in the reading comprehension assumes the important role of allowing the planning of more effective interventions, not only destined to develop the comprehension, but also to evaluate it. The evaluation, when accomplished from a productive theoretical referential, can inform a lot about the mental process and the authentic ability of the candidate's comprehension. The evaluative action includes exactly the understanding of the cognition process. When discussing underlying aspects of the reading comprehension evaluation, specifically in foreign language, in this case English language, the present research searched to investigate if the level or levels of textual comprehension reached by the subjects when reading a text were enough to execute the tasks of the test. The subjects were divided into two groups and submitted to two types of evaluations with different approaches: a standardized one, of multiple-choice type, in the molds of the test TOEFL and an alternative one, developed specially for this research, in the molds of Barret's taxonomy (1968). After each evaluation the subjects were asked to make a recount of the text previously read and to answer about the reading strategies used by them. The results report that, in the case of TOEFL, among the data from the sample, there was not a correlation among the levels of textual comprehension reached by the subjects and the result in the test. In other words, it was verified that through a particular reading behavior, that makes use of strategic reading, it is possible to answer the test satisfactorily, reaching a high score, without necessarily having reached a comprehension level that guarantees a deep representation of the text. Such fact highlights the importance of a reflection before taking decisions based on the tests' results.

Key- words: cognition, reading comprehension, evaluation

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

SUMÁRIO

INTRODUÇÃO

> [...] a relação entre o pensamento e a palavra não é uma coisa mas um processo, um movimento contínuo de vaivém do pensamento para a palavra, e vice-versa [...]. O pensamento não é simplesmente expresso em palavras; é por meio delas que ele passa a existir. (VYGOTSKY, 1993)

A relação feita por Vygotsky entre pensamento e linguagem ressalta a importância do significado como sendo um componente essencial da palavra e, ao mesmo tempo, um ato de pensamento. Sua percepção de que o significado de uma palavra já é, em si, uma generalização traz à tona a visão do processo no qual, através do significado da palavra, o pensamento e a fala se unem em pensamento verbal. Para Vygotsky, é no significado que se encontra a unidade das duas funções básicas da linguagem: o intercâmbio social e o pensamento generalizante. São os significados que vão propiciar a mediação simbólica entre o indivíduo e o mundo real, constituindo-se no filtro através do qual o indivíduo é capaz de compreender o mundo e agir sobre ele.

Estudar a leitura como processo de construção de sentidos se apresenta como um caminho para se prosseguir nas investigações acerca da cognição, estabelecendo-se então a importância da leitura para a cognição. Os fatores envolvidos no processo de construção do significado dos textos, ou seja, na compreensão leitora, são inúmeros, uma vez que essa, como processo mental que é, se constitui num fenômeno muito complexo. É devido a essa complexidade do processo de leitura que muitos pesquisadores vêm se dedicando a uma análise teórica em profundidade que, segundo Alliende e Condemarín (2005), facilite ao leitor detectar as fontes das dificuldades em compreensão e os modos de amenizá-las.

A tomada de consciência acerca dos fatores envolvidos na compreensão assume o importante papel de permitir o planejamento de atividades cada vez mais variadas, destinadas não apenas a desenvolver a compreensão, mas também a avaliá-la, por isso serão elencados no capítulo 1 alguns fatores cognitivos envolvidos nesse processo. Mas ocorre, então, a pergunta: por que é necessário avaliar a compreensão?

A busca pela avaliação, tanto formal quanto informal, é uma característica intrínseca ao homem que, enquanto ser cognoscitivo, monitora seu saber, tomando por base as ações avaliativas que recebe ao longo da vida. No caso específico da avaliação da compreensão leitora, como já foi dito, quanto mais se souber a respeito desse processo, mais se saberá sobre a cognição humana, assim a verificação de como se dá a compreensão fornece pistas importantes para melhor entender a cognição. A ação avaliativa abrange justamente a compreensão do processo de cognição.

Porém, a avaliação também cumpre outras funções que são, em parte, responsáveis pela concepção pejorativa inerente ao termo. A avaliação normativa, que, segundo Hoffman (2001, p.13), é "estática e frenadora", possui caráter classificatório e fundamentalmente "sentencivo". Esse tema das ações avaliativas e seus propósitos será abordado no capítulo 1. Serão ainda apresentados alguns exemplos de instrumentos de avaliação informal e formal utilizados no universo escolar, acadêmico e profissional.

Dentre os instrumentos apresentados, será destacado o TOEFL (*Test of English as a Foreign Language*), um teste de múltipla-escolha, internacionalmente aceito, aplicado para atestar o nível de compreensão leitora dos candidatos a vagas acadêmicas ou empresariais. Devido à sua importância e por consistir num dos principais parâmetros de avaliação da compreensão leitora em uma língua estrangeira, esse instrumento é objeto de estudo da presente investigação.

O objetivo proposto nesta pesquisa é verificar o grau de profundidade de compreensão obtido por meio de instrumentos como o TOEFL e sobretudo pela forma com que são aplicados. O que se quer com isso é olhar para a concepção de compreensão subjacente a tais testes tendo em vista seu impacto no sistema educacional.

Vários aspectos da avaliação serão enfocados, como a validação dos instrumentos, etapa importante no processo de elaboração e construção de testes de itens objetivos, uma vez que é mediante essa etapa que se define o grau de confiabilidade do instrumento. A preocupação que se tem é verificar, tendo em vista os vários aspectos cognitivos envolvidos na compreensão que devem ser considerados, se o teste dá conta dos objetivos a que se propõe, ou acaba priorizando uma determinada capacidade cognitiva como definidora de compreensão, em detrimento de outras.

A validade de um instrumento é definida em *A Dictionary of Reading and Related Terms* (HARRIS e HODGES, 1981, p. 276) como “a característica que determina que o teste mede o que se supõe que ele meça e/ou que ele possa prever o desempenho em outras medidas”[2].

No capítulo 3, são apresentados e discutidos os resultados da pesquisa. Nesse capítulo é analisada a abordagem do teste TOEFL, com base nos construtos teóricos apresentados no referencial e feita uma comparação entre os resultados obtidos no TOEFL e os obtidos num outro teste, desenvolvido para a pesquisa, com base na taxonomia de Barret, também apresentada no capítulo 1. Para concluir, algumas considerações finais.

---

[2] No original: *the characteristic of a test that determines that the test measures what it is supposed to measure and/or it can predict performance on other measures.*

1 ESTADO DA ARTE: COMPREENSÃO LEITORA UM MESMO OBJETO, VÁRIAS LENTES

> O resultado do encontro entre leitor e texto, que vem a ser o fenômeno da compreensão em leitura, não pode ser descrito, portanto, a partir de um único enfoque. (LEFFA,1996)

O avaliador de pedras preciosas analisa seu objeto sob diferentes ângulos, luzes e perspectivas. Assim é a proposta neste trabalho: analisar a questão da leitura, tendo como foco a compreensão e sua conseqüente avaliação, sob ângulos e graus variados à luz dos diversos olhares que vêm recebendo ao longo da história, através de uma breve revisão da literatura. Essa revisão pretende assim, traçar o quadro conceitual para a compreensão da pesquisa. Serão privilegiados definições e modelos que embasam o estudo da compreensão em leitura e sua conseqüente avaliação. Ela tem início com uma perspectiva histórica trazendo diversas concepções sobre a leitura e os elementos envolvidos no processo. Na seqüência é discutido o desenvolvimento das habilidade de leitura contemplando a necessidade de uma renovação da prática psicopedagógica no processo de aprendizagem. Na seção 1.2.2 é enfocada a compreensão em língua estrangeira. As sub-seções seguintes abordam os aspectos cognitivos da compreensão. Na seção 1.3 são revisitados alguns dos modelos existentes na literatura que visam a descrever o processo de leitura. Fechando a seção, é apresentado o aporte teórico que embasa a pesquisa.

Na seção 1.4, é abordado o tema da tradicional avaliação da compreensão leitora, enfocando, em especial, a avaliação da compreensão em língua estrangeira e os propósitos da mesma. São apresentados alguns instrumentos padronizados utilizados para medir a compreensão e é descrito o processo de validação dos mesmos. Nessa seção é destacado o

modelo de utilidade dos testes que serve como um importante critério para a atribuição de validade. Entre os instrumentos apresentados, o *Test Of English as a Foreign Language* (TOEFL), que consiste em um teste padronizado de avaliação da compreensão aceito internacionalmente como referência, receberá um enfoque especial por ser objeto de estudo da presente investigação. A seção termina com um olhar para o efeito que os testes de avaliação da compreensão em leitura exercem no sistema educacional e na sociedade como um todo.

1.1 A compreensão leitora sob diversos *olhares*

> A leitura, em seu sentido estrito, constitui um processo ativo de comunicação que leva o leitor a construir, intencionalmente, a partir da percepção de signos gráficos e da ajuda de dados não-visuais, uma substância de conteúdo equivalente àquela que o autor quis expressar, através de uma mensagem verbal escrita. (POERSCH e PORTO AMARAL, 1989, p. 133)

Conforme Leffa (1996), na epígrafe de abertura deste capítulo, o fenômeno da compreensão em leitura não pode ser descrito a partir de um único enfoque. Assim, será apresentado na seqüência um olhar caleidoscópico sobre o tema, bem como uma definição acerca dos construtos abordados aqui, ou seja, será definida a *lente* escolhida para observar o fenômeno cognitivo da compreensão em leitura.

O conceito de leitura vem sofrendo visíveis variações ao longo dos últimos anos. As definições mais antigas, anteriores a 1970, referiam-se à leitura não como um processo, mas como um produto que ocorria a partir da recepção e decodificação do *input* de linguagem, no caso, o texto.

Essa visão tem o leitor como um elemento passivo no processo de leitura. Segundo Ferreiro e Palaccio (1990), os pesquisadores que trabalhavam na área dos problemas de aprendizagem costumavam destacar a decodificação como o fator responsável pelas dificuldades de aprendizado. Conforme Carrell (1998), a partir de uma citação que resume bem a visão de leitura anterior a 1970, “a leitura [era] vista primeiramente como um processo

decodificador de reconstrução do sentido pretendido pelo autor por meio do reconhecimento de letras e palavras impressas."[3]

Novas perspectivas para a definição de leitura começam a surgir após 1970, e a decodificação passa a ser considerada como uma condição necessária porém não suficiente para a leitura.

Nos estudos de Halliday e Hasan (1989), por exemplo, é apresentado um novo enfoque da leitura de texto como processo que constitui-se no potencial de sentido que está a serviço de abordagens diferenciadas, em momentos também distintos, subjacentes às palavras, às sentenças, aos períodos, aos parágrafos.

Urquhart e Weir (1998) apresentam uma síntese feita por Urquhart em 1987 acerca das hipóteses mais comuns ainda existentes na época por trás da perspectiva pedagógica tradicional da compreensão:

> Hipótese 1: Há algo chamado de compreensão 'total' ou 'perfeita' de um texto. Hipótese 2: A leitura cuidadosa, que busca extrair uma perfeita compreensão, é superior a qualquer outro tipo de leitura [...] e é, na verdade, o único tipo de leitura que merece esse nome[4]

Entre os pesquisadores que comungavam de tal olhar, os autores citam Fry em 1963 e Sticht em 1984. Porém, de acordo com Urquhart e Weir (1998), independentemente de a 'compreensão perfeita' ser ou não uma meta exeqüível, deve-se rejeitar as afirmações de Fry e companheiros, de que esta seja equivalente a um resultado de 100 por cento em testes de compreensão. Os autores apontam que as tradicionais questões, freqüentemente de múltipla-escolha, que servem como um teste de compreensão, representam no máximo uma amostra da informação ganha com a leitura.

Os autores vêem ainda essa típica perspectiva pedagógica da compreensão como perigosa, podendo vir a causar um impacto no sistema educacional. Primeiramente por insistir na supremacia de um tipo de leitura em detrimento de todos os outros, ignorando os propósitos de leitura e tendo, como conseqüência, o menosprezo por outros comportamentos

---

[3] No original: *reading viewed primarily as a decoding process of reconstructing the author's intended meaning via recognizing the printed letters and words*. (p. 2)

[4] No original: *Assumption 1. There is such a thing as 'total' or 'perfect' comprehension of a text. Assumption2. Careful reading, which aims to extract perfect comprehension, is superior to any other kind of reading […] and is, in fact, the only kind of reading which deserves the name* (p.86).

de leitura que seriam perfeitamente aceitos dentro de dadas circunstâncias contextuais. Esse impacto será discutido com mais profundidade ao final deste capítulo.

Em segundo lugar, a perspectiva oferece perigo por ver o texto contendo uma quantidade finita de informações, acessíveis a todos os leitores. Segundo os autores, dentro dessa perspectiva, a informação estaria toda ‘na página’, o que acaba indo totalmente de encontro com a visão mais atual em que o leitor interage com o texto a fim de construir o sentido. Esse choque de visões é representado através da seguinte afirmação feita por Samuels e Kamil em 1988, citada em Urquhart e Weir (1998) :

> Assim ao contrário da sabedoria convencional que afirma que a compreensão é o processo de **retirar** o sentido de uma página, a compreensão é [...] o processo de **trazer** sentido a um texto[5].

As concepções mais recentes, como a que inicia essa seção, apresentam a leitura, ou a compreensão em leitura, como uma atividade mental de processamento e integração da informação e não como uma atitude passiva, de **extração** de sentido. A compreensão em leitura apresenta-se como uma constelação de capacidades e habilidades cognitivas inter-relacionadas, variando desde mapeamentos mecânicos a manipulações mais conceitualmente sofisticadas como raciocínio e inferência. Nessa nova concepção, a compreensão passa a ser vista como a base e não a conseqüência da leitura, como se costumava *ver*.

Segundo Smith (1989), imbuído de uma perspectiva vygotskyana sobre a relação pensamento-linguagem, não existe muita diferença entre ler e pensar; a leitura é um pensamento estimulado pela língua escrita, em que a atividade mental centra-se na compreensão de um texto escrito. Afirma ainda o autor que

> a leitura é uma atividade construtiva e criativa, tendo quatro características distintivas e fundamentais: é objetiva, é seletiva, é antecipatória e é baseada na compreensão, temas sobre as quais o leitor deve, claramente, exercer o controle. (SMITH, 1989, p.17)

1.1.1 O desenvolvimento das habilidades de leitura

---

[5] No original: *Thus, contrary to conventional wisdom, which states that comprehension is the process of getting meaning from a page, comprehension is […]the process of bringing meaning to a text* (p. 87) (grifo nosso).

Nunes e colaboradores (2003) apontam que foi apenas nas últimas duas décadas que os trabalhos de pesquisadores como Ferreiro e Teberosky (1985), entre outros, permitiram que se começasse a compreender os estágios que descrevem a evolução da leitura e escrita na criança. Os trabalhos anteriores, voltados para a análise dos processos de aquisição da leitura e da escrita, eram predominantemente pedagógicos, discutindo os métodos de ensino mais do que a aprendizagem.

Considerando a complexidade dos processos envolvidos na leitura proficiente, é compreensível que o aprendizado dessas habilidades imponha uma série de dificuldades ao aprendiz, fazendo-se necessário um olhar atento por parte dos educadores acerca dessas dificuldades.

A renovação da prática psicopedagógica também depende, segundo Nunes e colaboradores (2003), da renovação da concepção da leitura e escrita como atividades-meio e não exclusivamente como atividades-fim. Conforme os autores, “não se lê apenas para alcançar um decifrado, mas para conhecer o significado de um recado, **para aprender sobre um assunto** ou ainda como diversão” (op cit. p. 94 grifo nosso).

O aluno precisa *aprender* a *ler* para poder *ler* para *aprender*, tendo assim a leitura como atividade-meio que lhe servirá de dispositivo para representar mentalmente aquilo que lê num processo de construção de sentidos e integração com as informações que já dispunha em seu conhecimento prévio resultando na aprendizagem. Assim sendo, ler é compreender e, tomando emprestada a lente de Alliende e Condemarín (2005, p. 123), só “há compreensão quando, a partir do que já se conhece, se pode chegar a novos conhecimentos”.

Colomer e Camps (2002) apontam que, embora ler seja a base de quase todas as atividades que se realiza na escola, e a concepção de leitura como ato compreensivo seja aceita de um modo geral, a maioria das pesquisas sobre as atividades de leitura na escola, tais como as de Rocckwell (1982), Collins e Smith (1980) e Sole (1987), citadas pelos autores, demonstram que nelas não se ensina a entender os textos.

A simples decifração de palavras isoladas (atividade-fim), segundo Silveira (1995), ainda está muito longe da compreensão, pois tal atividade não atingiu ainda as perspectivas textual, discursiva ou pragmática. A identificação ou decifração de palavras é, sem dúvida,

como já dito, considerada uma condição necessária, mas não suficiente para se compreender um texto.

O ato de leitura implica dois movimentos complementares: primeiramente há que se desvendar um código, o que é possível após a aquisição de tal código, ou seja do alfabeto nas escritas alfabéticas; e em segundo lugar, há que se atribuir sentido ao que foi decodificado. Esses dois movimentos fazem a distinção entre alfabetização e letramento, conceito ainda recente na seara da educação.

Segundo Magda Soares (2001), ambas as ações envolvem as habilidades de leitura e escrita, todavia a alfabetização restringe-se a ter aprendido a ler e a escrever, não ultrapassando o significado de “levar à aquisição do alfabeto”, enquanto o letramento permite ao leitor fazer uso dessas duas habilidades e com elas interagir com o meio social em que vive.

> (...) um indivíduo alfabetizado não é necessariamente um indivíduo letrado; alfabetizado é aquele indivíduo que sabe ler e escrever; já o indivíduo letrado e, o indivíduo que vive em estado de letramento, é não só aquele que sabe ler e escrever, mas aquele que usa socialmente a leitura e a escrita, pratica a leitura e a escrita, responde adequadamente às demandas sociais de leitura e de escrita. (SOARES, 2001, p. 39-40)

Assim pode ser observada a diferenciação entre um processo de aquisição da língua, que é finito, e um processo de desenvolvimento da língua, que é contínuo. A união desses processos exerce um papel fundamental na melhoria da qualidade de vida dos indivíduos.

Conforme Silveira (1995), uma vez examinada a prática escolar da leitura ao longo das oito séries do ensino fundamental, vê-se, primeiramente que, de um modo geral, a escola não se preocupa com o ***uso social da leitura***, ou seja, não são levados em consideração os vários usos da leitura na vida real das pessoas enquanto cidadãos inseridos numa comunidade letrada. A escola, conforme a autora, não estaria, portanto, dando conta daquelas três funções básicas da leitura, mencionadas por Geraldi (1984), quais sejam***:*** a leitura-deleite ou fruição, a leitura para aquisição de informações e a leitura para estudo e trabalho. A prática escolar da leitura ainda é estereotipada e unidirecional, tendo como única fonte de textos o livro didático, cujas atividades e tipos de texto (literários, na grande

maioria) são repetitivos e usados de forma desvirtuada de sua verdadeira função (Silveira,1995).

Ao oportunizar a prática, a reflexão e ação em sala de aula, trabalhando com diversos suportes, tipos e gêneros textuais nas atividades de leitura com as crianças, a prática escolar viabilizaria a emergência de “imensos tesouros de um sentido potencial” (BAKHTIN, 1997, p. 69). O trabalho com o gênero da propaganda, por exemplo, que é uma prática pouco difundida na educação infantil, é uma forma possível de se aproximar a leitura de mundo com a leitura da palavra, uma vez que, segundo Marcuschi (2002), os gêneros sociais são o reflexo de estruturas sociais recorrentes e típicas de cada cultura.

Conforme o autor, tendo em vista que todos os textos se manifestam sempre num ou noutro gênero textual, um maior conhecimento por parte dos leitores acerca do funcionamento dos diversos gêneros textuais é importante tanto para a produção como para a compreensão de textos. Os gêneros são modelos comunicativos, servindo, muitas vezes, para criar uma expectativa no interlocutor e prepará-lo para uma determinada reação. Segundo Bakhtin (1997, p. 71), “os gêneros operam prospectivamente, abrindo o caminho da compreensão”.

Até este momento, ao se falar em compreensão, está se pensando no fenômeno em si e de um modo generalizador, não vinculado a uma determinada língua. Essa postura é defendida por Goodman (1987), para quem o processamento de leitura é o mesmo para todas as línguas, uma vez que o esforço que subjaz o trabalho do leitor em construir o sentido do texto requer a utilização de estratégias e habilidades cognitivas independentemente do idioma que utiliza. Um contraponto possível acerca desse olhar de Goodman seria em relação ao domínio da língua. Para que o processamento de leitura seja o mesmo para mais de uma língua, o domínio dessas línguas deveria ser semelhante ao domínio na língua materna. A seguir será enfocada a compreensão em leitura especificamente em língua estrangeira, uma vez que essa é o alvo da presente investigação.

1.1.2 A compreensão leitora em língua estrangeira

No caso específico de uma língua estrangeira, também se pode perceber diferentes concepções ao longo da história. A leitura era vista como um adjunto para as habilidades de linguagem oral. Segundo Carrell (1998), o método audiolingual influenciou a concepção da primazia do áudio sobre a leitura e da fala sobre a escrita. Outra influência, segundo a autora, para a promulgação e implementação da perspectiva de decodificação na leitura em língua estrangeira foi a importância dada por estruturalistas para as relações grafema-fonema.

A compreensão em leitura em segunda língua, doravante L2[6], é, conforme Koda (2005), tradicionalmente vista como um aspecto integrante da proficiência da língua e, como conseqüência, a posterior avaliação da compreensão leitora em L2 reflete a proficiência em L2. A autora atenta para o fato de que apenas recentemente o elemento central na proficiência de L2 mudou de um conhecimento lingüístico discreto e descontextualizado para uma competência comunicativa. Mudanças paralelas refletindo essa alternância de perspectiva também ocorreram na avaliação da compreensão leitora em L2.

Aebersold e Field (1998), investigando os fatores que influenciam a leitura numa L2, apresentam as quatro áreas envolvidas no processo de compreensão: competência gramatical, sócio-lingüística, discursiva e estratégica. Com a mudança de percepção acerca do elemento central na proficiência de L2, a atenção volta-se para fatores relativos à intenção comunicativa e à construção de sentido.

Além da importância dada à intenção comunicativa e à construção de sentido, outra importante percepção ganha luz: ler em uma língua estrangeira envolve duas ou mais línguas, por isso as análises deveriam ser translingüísticas, explorando ambas as características de língua materna (L1) e L2 como possíveis fontes de diferenças individuais no desenvolvimento da leitura em L2 (KODA, 2005).

Uma proposta de descrição do processo da compreensão deve levar em conta ao menos alguns aspectos cognitivos essenciais como o autor, o texto, o leitor e as circunstâncias em que se dá o encontro. A próxima seção contempla alguns desses aspectos.

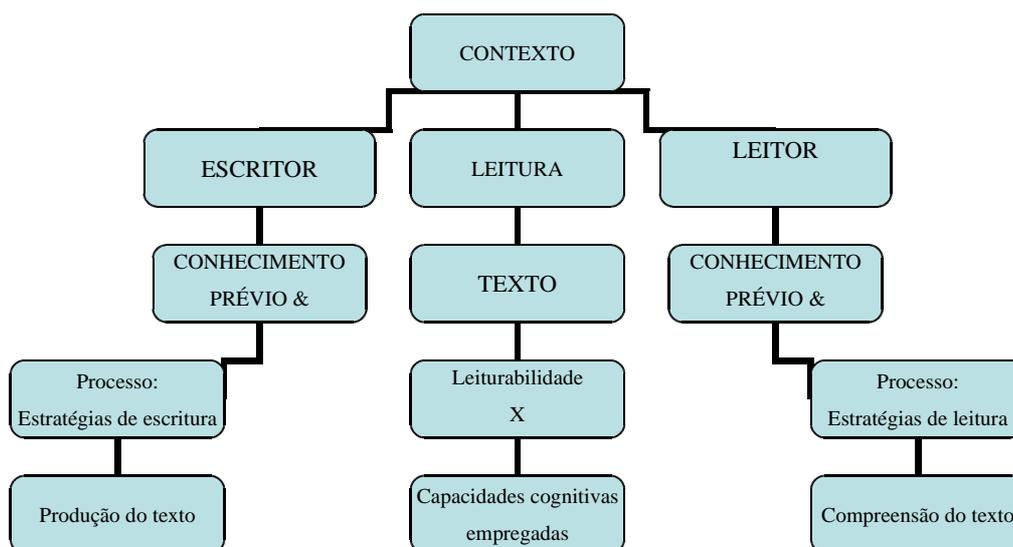
[6] Embora exista distinção na literatura entre L2 e língua estrangeira, será utilizado aqui o termo L2 para fazer referência à língua Inglesa, que vem a ser uma língua estrangeira para os sujeitos da pesquisa falantes de português brasileiro.

1.1.3 Aspectos cognitivos da compreensão leitora

Como visto, a concepção de compreensão/leitura varia conforme a lente que se utiliza para analisá-la, por isso a intenção é deixar claro para o leitor desta dissertação com que *grau* são vistos aqui os aspectos cognitivos presentes na compreensão textual. Assim, partimos de Weinstein e Mayer (1986 apud KODA 2005), para quem os aspectos ou processos cognitivos referem-se a um grande grupo de ações que auxiliam a governar comportamento, emoção, motivação, comunicação, atenção e compreensão, até chegarmos a Damásio (2003) para quem

> não parece sensato excluir as emoções e os sentimentos de qualquer concepção geral da mente, muito embora seja exatamente o que vários estudos científicos e respeitáveis fazem quando separam as emoções e os sentimentos dos tratamentos dos sistemas cognitivos[...] as emoções e os sentimentos são considerados entidades diáfanas, incapazes de compartilhar o palco com o conteúdo palpável dos pensamentos, que não obstante qualificam. (DAMÁSIO, 2003, p. 189)

Desse modo, ao se falar em aspectos cognitivos da leitura, se está considerando os sentimentos, uma vez que esses, conforme Damásio (2003, p. 190), "são tão cognitivos como qualquer outra imagem perceptual e tão dependentes do córtex cerebral como qualquer outra imagem".

Assim, de uma forma panorâmica, estão envolvidos no processo de compreensão leitora, seja essa em L1 ou L2, além dos aspectos derivados do leitor, os derivados do texto e os do autor. No Quadro 1, é apresentado um esquema que busca elencar alguns dos diversos componentes envolvidos na leitura, bem como suas possíveis intersecções, embora não seja possível estabelecer precisamente uma hierarquização desses elementos.

Quadro 1 – Quadro dos elementos envolvidos na leitura
Fonte: adaptação (GABRIEL, 2005, p. 170)

O quadro acima será explicado nas três sub-seções seguintes, dando continuidade à revisão da literatura. Essa explicação será dividida em aspectos da compreensão derivados do autor, do texto e do leitor, respectivamente. Segundo Leffa (1996), historicamente, o estudo da compreensão de leitura tem se caracterizado pela predominância de um ou outro extremo do processo, enfatizando ora o texto, ora o leitor, como fator essencial da compreensão. Cada um desses enfoques pressupõe uma explicação diferente para os fatores que intervêm na compreensão.

Quando se privilegia o texto, por exemplo, pressupõe-se que a melhoria na compreensão depende de qualidades intrínsecas do texto e que, na medida em que se modificam essas qualidades, está-se modificando os níveis de compreensão do leitor. Quando se privilegia o leitor, pressupõe-se que a compreensão do texto aumenta na medida em que se desenvolve no leitor as habilidades gerais da leitura (LEFFA, 1996).

Uma vez que esse levantamento histórico sugere uma divisão entre texto e leitor, partimos, para fins didáticos, dessa dicotomia. Conforme o autor, as implicações pedagógicas desse levantamento são óbvias. Na medida em que se é capaz de destacar os fatores mais e menos importantes para a compreensão de leitura, se está dando pistas ao professor sobre como atuar com seus alunos leitores. Seguem então os elementos.

1.1.3.1 Aspectos cognitivos da compreensão leitora derivados do escritor

O escritor, lançando mão de um conjunto de códigos que maneja e de seu patrimônio cultural e de conhecimento prévio, produz uma mensagem (texto) num determinado contexto, que poderia ser chamado de "circunstância da escrita". Algumas vezes, o escritor tem em mente, no momento da escrita, um leitor ideal e escreve basicamente para esse leitor, levando em conta os códigos e conhecimentos prévios de seu receptor para que a mensagem possa ser adequadamente interpretada. Um exemplo desse tipo de atitude pode ser observado na propaganda de produtos caros, como tênis, em que o texto aparece apenas em inglês, prevendo que o consumidor potencial do produto domina essa L2.

O conhecimento das circunstâncias da escrita pode ser fundamental para a compreensão de um texto, pois, em alguns casos, saber o lugar e a época em que o texto foi escrito, ou a idade, cargo ou crença política que o autor tinha no momento da escritura pode viabilizar a compreensão. Porém, em muitos outros casos, o texto pode ser compreendido prescindindo-se dos fatores derivados do autor.

1.1.3.2 Aspectos cognitivos da compreensão leitora derivados do texto

Conforme Alliende e Condemarín (2005), depois de criado, o texto passa a ter uma existência própria, independente do autor. Os principais fatores do texto que incidem na compreensão leitora podem ser divididos, ainda segundo os autores, em físicos, lingüísticos e referenciais ou de conteúdo. O conjunto desses fatores pode determinar a legibilidade ou leiturabilidade de um texto.

A legibilidade física é constituída pelos fatores materiais do texto, tais como tamanho ou formato das letras e/ou do texto ou qualidade de impressão, ou qualidade da tela (no caso de material no computador) que influem na possibilidade de o texto ser compreendido. Porém, não é critério único para a impossibilidade da leitura, apenas torna a compreensão mais difícil. Esse tipo de legibilidade tem hoje grande importância, dada a variedade de suportes em que os textos são lidos: papel, tela de computador, projeções, entre outros.

A legibilidade lingüística, ou critério de leiturabilidade, relaciona-se aos aspectos lingüísticos do texto. Um texto com um conteúdo lexical muito inacessível para o leitor, com estruturas morfossintáticas muito complexas, ou com elementos dêiticos em demasia tem sua compreensão dificultada.

A legibilidade referencial de um texto diz respeito aos conteúdos do mesmo e influi em sua compreensão. Segundo Alliende e Condemarín (2005), qualquer referente pode ser incluído num texto, mas essa inclusão não se dá de forma direta, passando antes por um processo de transformação: transforma-se numa entidade semântica (a generalização vygotskyana) e em parte de uma comunicação. Se a compreensão leitora versasse sobre o conteúdo sem a transformação em entidade semântica, o leitor teria que saber absolutamente tudo sobre tal conteúdo.

Ao serem transformados em entidades semânticas, os referentes passam a ter características diferentes das do objeto representado e, ao serem incluídos no texto, passam a ser objeto da interação dos outros elementos textuais, estabelecendo com eles uma infindável quantidade de relações. Um meio para se obter adequadas categorias de análise para um determinado texto poderia ser determinar os diferentes códigos que podem ser descobertos nele e aplicá-los aos referentes (ALLIENDE E CONDEMARÍN, 2005).

Os autores apontam que grande parte das provas de compreensão leitora aplicadas nos estabelecimentos educacionais versa sobre os conteúdos dos textos. E, uma vez que esses conteúdos não podem ser determinados previamente, apela-se para certas categorias gerais supostamente aplicáveis a qualquer texto: determinação de idéias principais e secundárias; captação de detalhes; seqüências; relações de causa-efeito; captação de atitudes, emoções, sentimentos, interesses, ideais, características pessoais; relações de semelhança, oposição, inclusão, exclusão, entre outras.

Assim, o conteúdo dos textos passa a ser um dos principais fatores que incidem sobre a compreensão, particularmente nos níveis avançados. A compreensão dos referentes não depende tanto do conhecimento do objeto representado em si, mas de suas características como unidade semântica que se inclui e se articula com o texto, formando conjuntos que podem ser incluídos em esquemas significativos que permitem a compreensão em qualquer nível.

Outra forma de incidência do conteúdo dos textos sobre a compreensão é que é ele quem define o critério de profundidade, apontado em Gabriel (1998), segundo o qual é possível analisar os diferentes níveis da compreensão, uma vez que nem sempre ela se dá em plenitude, não se configurando como fenômeno *tudo-ou-nada*.

O critério de profundidade relaciona-se, assim, aos níveis em que se constrói o sentido do texto. O conteúdo explícito corresponde ao que está aparente e claramente expresso nas linhas do texto. O conteúdo implícito corresponde ao sentido que deve ser *lido* para favorecer a compreensão, embora não esteja *escrito*. É o conteúdo expresso nas entrelinhas do texto: as elipses, as ironias e inferências. Para acessar esse conteúdo, o leitor precisa lançar mão de uma série de estratégias que serão abordadas na próxima seção. Por último, o conteúdo metaplícito é aquele que é construído unicamente mediante conhecimento da circunstância da escrita, ou seja, do contexto de escrita e, para isso, é necessário um conhecimento extratextual (POERSCH, 1991). Esse conhecimento é, em grande parte, o responsável pela diversidade de interpretações, uma vez que ele pode variar de leitor para leitor.

1.1.3.3 Aspectos cognitivos da compreensão leitora derivados do leitor

Os aspectos derivados do leitor são inúmeros, porém por questões de tempo e espaço, serão enfocados a seguir alguns dos elementos mais recorrentes no processo de compreensão leitora e que vêm sendo citados mais continuamente na literatura. O primeiro refere-se às estratégias de leitura adotadas pelo leitor no momento da leitura. Conforme Koda (2005),

> leitores competentes ajustam continuamente seus comportamentos de leitura para acomodar dificuldades com textos, realização de tarefas e outras variáveis contextuais. Monitorando cuidadosamente seu processo de leitura, eles tomam providências imediatas quando encontram problemas de compreensão. Conscientes de seus próprios recursos cognitivos e lingüísticos eles são capazes de dirigir sua atenção para as pistas apropriadas ao antecipar, organizar e reter informação textual. Esses e outros comportamentos semelhantes, separando leitores competentes dos menos competentes, caracterizam o que é referido como leitura estratégica[7]. (KODA, 2005)

[7] No original: *Accomplished readers continuously adjust their reading behaviors to accommodate text difficulty, task demands, and other contextual variables. Monitoring their reading process carefully, they take immediate steps when encountering comprehension problems. Aware of their own cognitive and linguistic resources, they are capable of directing their attention to the appropriate clues in anticipating, organizing, and retaining text*

O escritor, baseando-se no fato de que o leitor dará uma série de contribuições ao texto, processando e integrando informações a partir de seu conhecimento prévio (memória de longo prazo), deixa uma série de lacunas e informações incompletas para o leitor preencher. Segundo Alliende e Condemarín (2005), muitas vezes uma única linha de texto desencadeia uma enorme quantidade de processos mentais no leitor.

Esse processamento avança em duas direções: de baixo para cima, o chamado processamento ascendente ou *bottom-up*, e de cima para baixo, o chamado processamento descendente ou *top-down*.

No primeiro tipo, o leitor começa por fixar-se nos níveis inferiores do texto, tais como os sinais gráficos e as palavras (microestruturas) para formar, mediante o acesso à representação fonológica de cada grafema, sucessivamente, as diferentes unidades lingüísticas até chegar aos níveis superiores da frase e do texto, ou seja, a macroestrutura. No segundo, o movimento é inverso, atuando a partir da mente do leitor para o texto. Esse processamento baseia-se em fatores tais como o objetivo do leitor ao ler, seu conhecimento prévio, as hipóteses e inferências contínuas por ele levantadas ao longo da leitura e a avaliação da pertinência dessas inferências (COLOMER e CAMPS 2002).

Conforme Alliende e Condemarín (2005), em função desse duplo processo, a compreensão de um texto depende, em parte, dos esquemas do leitor. Cada vez que um leitor vivencia algo, seja direta ou indiretamente, mediante a leitura, relato ou vivência, ele aplica um esquema a essa experiência. O esquema inicial pode, segundo os autores, se alterar profundamente, ou se confirmar e tornar-se mais preciso.

É por meio desses esquemas que o leitor pode lançar mão de outra estratégia para completar as lacunas deixadas propositalmente pelo escritor: a operação de inferência. Essa estratégia procedimental é intensamente utilizada no processo de leitura. Segundo Kato (1995), a inferência construtiva é aquela que cria significados a partir de pistas contextuais. A utilização razoavelmente consciente dessa estratégia leva o leitor a prever que poderá fazer

---

information. These and similar behaviors, separating accomplished and less-accomplished readers, essentially characterize what is referred to as strategic reading (p.204).

inferências mal sucedidas, as quais serão refutadas, o que o faz desenvolver, paralelamente, procedimentos para a auto correção, monitorando sua compreensão.

Outras estratégias comumente utilizadas pelos leitores serão apresentadas na próxima seção associadas aos propósitos de leitura. O propósito de leitura em relação ao texto que está sendo processado constitui-se num outro aspecto cognitivo derivado do leitor que incide sobre a compreensão.

Cada propósito de leitura desencadeia uma forma diferente de processar a informação textual bem como o emprego de diferentes habilidades e estratégias. Para ilustrar tal posicionamento, Koda (2005) apresenta algumas situações de leitura em que o verbo *ler* poderia ser substituído por outros verbos (entre parênteses) que designariam o propósito de leitura daquela situação, bem como o modo de processá-la: *ler* (espiar) a agenda no escritório para confirmar o horário de uma reunião; *ler* (escanear) o manual de usuário para descobrir como instalar um aparelho de DVD; *ler* (folhear) o último número de uma revista popular para saber o paradeiro das celebridades; *ler* (examinar) um artigo de jornal para adquirir informações sobre a tecnologia de modificação de genes, ou *ler* (estudar) um livro de biologia para se preparar para uma prova.

Assim, à medida que o leitor identifica seu propósito de leitura, ele pode ajustar seu modo de processar o texto, lançando mão de estratégias cognitivas específicas para aquele processamento. No Quadro 2 são apresentadas algumas das estratégias e habilidades cognitivas mais recorrentes no processamento de textos associadas ao propósito de leitura correspondente.

Habilidades/estratégias	**Propósitos**
Leitura rápida, exploratória (*Skimming)*	Estabelecer um sentido geral Estabelecer a estrutura macroproposicional
Leitura de busca (*Search reading*)	Localizar informações relevantes para leitura cuidadosa
Leitura cuidadosa, varredura (*Scanning)*	Localizar informações específicas
Leitura cuidadosa de idéias principais explicitamente declaradas	Estabelecer a macroestrutura Ler para entender e aprender

Leitura cuidadosa de idéias principais implicitamente apontadas: inferir sentido proposicional	Descobrir a intenção do autor Entender a atitude do autor em relação ao tópico Identificar o destinatário Distinguir fato de ficção
Leitura cuidadosa de sentido relacionado ao texto: inferir sentido pragmático	Aplicar idéias principais a outros contextos Avaliar um ponto de vista Expressar opiniões próprias a respeito do assunto.

Quadro 2 –Habilidades e estratégias associadas aos propósitos de leitura
Fonte: adaptação (URQUHART e WEIR, 1998, p 185).

Outros aspectos cognitivos derivados do leitor que incidem sobre a compreensão do texto são o conhecimento prévio, ou seja a bagagem que o leitor traz, e a memória. No olhar de Schank e Abelson (1977)

> as pessoas necessitam de uma grande quantidade de conhecimentos para poder compreender. [A compreensão é] um processo pelo qual as pessoas relacionam o que vêem ou ouvem [ou lêem] com grupos de ações pré-armazenadas que experimentaram previamente. [...] A nova informação é entendida nos termos da antiga. (SCHANK e ABELSON, 1977, p. 54)

O esforço do leitor diante de um texto pode ser comparado ao trabalho do cortador de pedras da célebre fábula de Jacob Riis, em que, após 100 marteladas sem resultado algum, na centésima primeira vê a pedra se partir em pedacinhos e o cortador sabe que não foi aquela martelada apenas, mas todas as que a antecederam foram responsáveis pelo feito. Assim é o trabalho do leitor durante o ato de leitura: “martelando” continuamente na tentativa de integrar, conforme a citação de Schank e Abelson (1977), cada nova proposição trazida pelo texto com as já processadas anteriormente, que agora fazem parte de seu conhecimento prévio, ou seja, do conjunto de suas memórias (IZQUIERDO, 2002).

O conhecimento prévio se caracteriza pela bagagem de experiências que o leitor traz consigo para o trabalho com o texto. A própria experiência anterior de leitura influi no modo pelo qual o leitor processa o texto. Quanto mais prática de leitura, melhor o leitor lida com os fatores derivados do texto que determinam a leiturabilidade do mesmo; mais o leitor amplia seu domínio sobre assuntos diversos, alargando continuamente esse conhecimento.

O conhecimento prévio na compreensão em leitura assume um papel primordial. Conforme Kintsch (1998), em seu modelo de construção-integração, para haver a compreensão é necessário que haja a) um processo de construção no qual uma base textual é construída a partir da informação lingüística recebida (input) bem como da base de conhecimento do leitor (conhecimento de mundo), e b) uma fase de integração na qual essa base textual é integrada em um todo coerente.

A compreensão de um texto requer, assim, a ativação do conhecimento de mundo ou prévio do sujeito, que é, conforme Kleiman (1989, p. 21):

> O conhecimento parcial, estruturado que temos na memória sobre assuntos, situações, eventos (...). Para haver compreensão, durante a leitura, aquela parte de nosso conhecimento de mundo que é relevante à leitura do texto deve estar ativa num nível ciente.

A tarefa de integrar novas e antigas informações é executada pela memória de trabalho, a qual, segundo Izquierdo (2002; 2004), possui capacidade tanto de armazenamento temporário ativo, que persiste por alguns minutos ou segundos, quanto de processamento. É nesse ponto que consiste a relação entre conhecimento prévio e memória de trabalho que será discutida a seguir.

A memória, que é, conforme Izquierdo (2004), responsável pela aquisição, conservação e evocação de informações, pode ser classificada sob os critérios de duração e conteúdo. Quanto à sua duração, a memória divide-se em imediata, curta e longa. Quanto ao conteúdo, ela se divide em memória de trabalho (que não deixa arquivos permanentes), memória declarativa (que pode se subdividir em semântica e episódica) e memória procedural.

Desde a decodificação já se percebe a influência da memória na leitura, uma vez que o reconhecimento do próprio texto só é possível porque o código já foi aprendido e está na memória. Conforme Koda (2005), uma vez extraída do texto impresso, a informação lexical deve ser agrupada em unidades de sentido maiores como frases, orações e parágrafos. A memória de trabalho desempenha um papel central nesse processo.

Além do acesso lexical, virtualmente todas as operações na leitura passam pela memória. As altas correlações entre a capacidade de memória de trabalho e a performance em leitura (Daneman & Carpenter, 1980) são um forte indício de que tal capacidade é uma das principais fontes das diferenças individuais na habilidade de compreensão leitora.

Miyake e Friedman (1998) definem memória de trabalho como "uma arena computacional ou espaço de trabalho, abastecido por recursos cognitivos limitados organizáveis flexivelmente, ou uma ativação que suporte tanto a execução de computações simbólicas (código) como a manutenção dos produtos intermediários gerados por essas computações"[8]. Esse enfoque ilumina a natureza dinâmica da memória de trabalho, oferecendo um contraste em relação à concepção tradicional de memória que atribui a ela a função passiva de estocagem (KODA, 2005).

Assim, conforme sugerem inúmeras pesquisas, a capacidade de memória de trabalho é particularmente importante nas tarefas cognitivas complexas que necessitam de operações de estágios múltiplos como resolução de ambigüidades, análise sintática e geração de inferência, entre outras estratégias de leitura empregadas pelo leitor na tarefa de compreensão.

Porém, segundo autores como Chistiansen e MacDonald (2002), diferenças individuais na compreensão não se originam de variações em uma capacidade de memória de trabalho isolada; ao invés disso, elas emergem de uma interação de fatores biológicos e experiência de linguagem. Essa alternativa, conforme os autores, é debatida para gerar uma contabilização superior de resultados de compreensão previamente atribuídos a uma capacidade de memória de trabalho separada.

Segundo Koda (2005), especial atenção deve ser dirigida aos testes utilizados para medir essa capacidade da memória de trabalho, pois as tarefas utilizadas nesses instrumentos às vezes investigam habilidades semelhantes, se não idênticas, subjacentes à leitura. O mesmo cuidado deve ser direcionado para os testes de itens objetivos, utilizados para medir a compreensão em leitura, que acabam algumas vezes por medir a memória de trabalho isoladamente.

[8] No original: *a computational arena or workspace, fueled by flexibly deployable, limited cognitive resources, or activation that supports both the execution of various symbolic computations and the maintenance of intermediate products generated by these computations". (p. 341)*

1.2 Modelos de leitura de textos e aporte teórico da pesquisa

Esta pesquisa procura investigar, sob a ótica cognitiva, a compreensão em leitura, especificamente em língua estrangeira, a partir da forma como se dá sua conseqüente avaliação. Nos parágrafos seguintes será feita uma breve referência a alguns dos modelos de leitura de texto presentes na literatura. Ao final desta seção se buscará demonstrar a correlação entre o paradigma que emoldura esta pesquisa e o modelo de leitura de texto escolhido aqui como lente para observar o fenômeno da compreensão.

Após os processos mais primários de leitura, como identificação e extração do significado de palavras individuais, como já dito, outros processos operam em diferentes graus de profundidade, agindo no nível da frase ou sentença, além de tratarem da organização global ou da estrutura temática do texto como um todo (SALLES, 2001). Vários modelos ao longo da história vêm tentando, cada qual através de sua lente própria, descrever esses processos.

É possível perceber, através de um olhar diacrônico sobre a literatura, que cada um dos modelos teóricos que surge com a proposta de descrever os processos de leitura acaba por negligenciar algum aspecto. E o modelo sucessor enfatiza exatamente aquele aspecto não considerado por seu predecessor, deixando de contemplar outros aspectos que serão abordados por seus sucessores e assim consecutivamente. Essa sucessão de modelos parciais denota o grau de complexidade do fenômeno em questão e os avanços graduais que a pesquisa em leitura vem fazendo ao longo de sua história.

Logo, a "dança" das sucessões de modelos teóricos pode ser vista como a mola propulsora que impulsiona o avanço da pesquisa em qualquer área. Se fosse utilizada uma imagem para esse olhar ela seria a de um "espiral: uma curva plana gerada por um ponto móvel que gira em torno de um ponto fixo, ao mesmo tempo em que dele se afasta ou se aproxima segundo uma lei determinada"[9]. Também é interessante observar o movimento circular que os próprios pesquisadores fazem, retomando as pesquisas anteriores, porém com um novo olhar.

[9] Definição encontrada em Aurélio, séc XXI. :

Grande parte dos modelos teóricos, como o de Graesser (1981), Gough (1972), Davies (1968) e Clymer (1968), que serão brevemente descritos a seguir, entre outros, enfocava “o quê” da leitura: “que” estratégias um leitor precisa para processar o texto; “que” tarefas demonstram tais estratégias; “que” tipos de compreensão existem e assim por diante. Em contraste, surgem outros modelos como os de Kintsch e van Dijk (1978) e Carpenter and Just (1988), que olham para o “como” da leitura: buscam investigar “como” se dá o processo de leitura; “como” ensiná-lo. Essa dicotomia se deve, em parte, à lente escolhida por cada modelo para explicar o processo de leitura. Enquanto alguns se apóiam substancialmente em perspectivas pedagógicas, outros tomam por base a psicologia ou a ciência cognitiva.

Uma observação detalhada desses modelos teóricos revela uma grande quantidade de variação existente entre eles com diversos níveis de descrição sendo enfocados. A seguir será apresentado, ainda que brevemente, um recorte com alguns dos modelos existentes.

Graesser (1981) descreveu seis fontes básicas de conhecimento envolvidas na compreensão textual: lingüística, retórica, causal, intencional, espacial e a de papéis, personalidades e objetos. Mackworth (1972), em contraste, tendo adotado outra abordagem, optou por focalizar o processo que ocorre durante a leitura, desenvolvendo um modelo de tarefa de leitura, que era essencialmente um modelo de processamento de informações.

Outros conhecidos pesquisadores também optaram por enfocar mais o processo de leitura em si do que o conhecimento necessário para a leitura. Um dos modelos mais detalhados, na interpretação de Moorman e Ram (1994), foi desenvolvido por Gough em 1972, e surgiu como um modelo de “um segundo de leitura”, o qual foi projetado para descrever todos os processos que perfazem um segundo de leitura oral.

Outro nível de descrição também enfocado por pesquisadores foi o modelo acerca das habilidades ou estratégias necessárias para a leitura proficiente. Davies (1968) identificou cinco habilidades necessárias para a leitura: memória para o significado das palavras; realizar inferências a partir do conteúdo; seguir a estrutura do texto; entender e/ou reconhecer o propósito, atitude, tom e estado de espírito do autor; ser capaz de responder questões tanto

explicitamente como em paráfrases. Tais habilidades estão em um nível intermediário entre conhecimento e processo e incluem elementos de ambos (MOORMAN e RAM, 1994).

Vários outros modelos buscando explicar a compreensão leitora estão disponíveis na literatura. Kingston (1961) apresenta um modelo descrito por ele mesmo como experimental, no qual o autor inclui nível de abstração, experiências, nível desenvolvimental e símbolos lingüísticos como variáveis. Com esse modelo, o autor fornece uma série de postulados para futuros modelos. Cleland (1965) também ofereceu um modelo de compreensão no qual ele especifica alguns dos principais processos constituintes da leitura.

Uma visão bastante extensa da constituição da compreensão leitora é revelada por meio da Taxonomia de Barret, apresentada em um artigo não publicado, mas que foi trazido a público, com a permissão do autor, por Clymer (1968) (ANEXO C). Nesse artigo, segundo Clymer (1968), Barret identifica duas concepções errôneas enfrentadas por alguns professores ao ensinarem a compreensão. “Essas concepções são: (a) considerar a compreensão uma habilidade única e (b) pressupor que a compreensão contém tantas habilidades separadas a ponto de não serem controláveis[10]” (CLYMER, 1968, p. 17). A fim de desenvolver um meio compreensível e manejável de ensinar a compreensão, Barret lança sua Taxonomia das Dimensões Afetivas e Cognitivas da Compreensão Leitora, que será apresentada com mais detalhes na seção 1.4.3.2 sobre a avaliação da compreensão[11].

A compreensão em leitura, na forma em que é tratada na taxonomia, é dividida em cinco grandes categorias ou níveis de habilidade: (a) compreensão literal; (b) reorganização; (c) compreensão inferencial; (d) avaliação e (e) apreciação. Dentro de cada uma dessas categorias, Barret fornece exemplos de tarefas específicas, no formato de propósitos de leitura, os quais o professor pode utilizar em sua prática em sala de aula. As categorias foram ordenadas de modo a se deslocarem das mais fáceis para as mais difíceis em termos das solicitações demandadas em cada categoria.

Esse sistema pode ter várias aplicações úteis para os professores. Primeiramente, a taxonomia pode ser utilizada como base para o desenvolvimento de propósitos e questões

[10] No original: *These misconceptions are: (a) considering comprehension a single unitary skill and (b) assuming that comprehension contains so many separate skills as to be unmanageable* (p. 17).
[11] Foi mantida aqui, como forma de preservar a produção original, a separação feita por Barret entre cognitivo e afetivo, embora comunguemos do olhar de Damásio para quem o afetivo faz parte do cognitivo.

para guiar a leitura. Em segundo lugar, os professores podem aplicar a taxonomia nas questões oferecidas pelos manuais para determinar que tipo de compreensão está sendo enfatizada naquela seleção (CLYMER, 1968).

Um outro modelo de compreensão, que exerceu forte influência nos estudos acerca da leitura, foi o resultado da pesquisa desenvolvida por Carpenter e Just (1988). Nesse modelo, os autores optaram por olhar para a leitura através da lente da memória de trabalho do leitor e de suas limitações. Buscaram investigar como essas limitações influenciam a leitura. Porém, como em todo processo de pesquisa científica, surge uma crítica específica a esse modelo. Segundo Christiansen e MacDonald (2002), a distinção comumente feita entre 'tarefas de processamento de linguagem e tarefas de memória de trabalho lingüística' é uma distinção artificial, e ambas as tarefas são simplesmente medidas diferentes da habilidade do processamento de linguagem.

Os autores baseiam suas asserções em uma abordagem conexionista para a compreensão da linguagem, que trata as distinções tradicionais entre *conhecimento lingüístico*, *processamento lingüístico* e *capacidade de processamento* de uma forma muito diferente. Segundo esses autores, as alegações de Just e Carpenter (1992) e Waters e Caplan (1996) estão baseadas em arquiteturas de processamento mais simbólicas. Como tudo depende da lente escolhida para se olhar para um dado objeto, a escolha da arquitetura de processamento tem um efeito direto nas asserções sobre o papel da memória de trabalho na compreensão da linguagem.

O paradigma conexionista (PDP, isto é processamento distribuído em paralelo), apoiado nas descobertas neurocientíficas, como a fisiologia do sistema nervoso central, tem como princípio fundamental o de que o conhecimento humano está armazenado em engramas, ou redes de conexão interneuroniais: sempre que os neurônios de uma rede forem ativados ao mesmo tempo, a conexão entre eles se tornará mais forte e a rede ficará "marcada". Cada vez que um estímulo (*input*) ativar um dos neurônios da rede, toda a rede será ativada (CHIELE, 2004). No caso da memória de trabalho, por exemplo, os neurônios do córtex pré-frontal se ativam em resposta às experiências momentâneas, e sua estimulação dura enquanto durar a experiência (IZQUIERDO, 2004).

O processamento distribuído em paralelo permite a ocorrência simultânea do reinstanciamento de várias informações, sua manipulação e mesmo o processamento concomitante de vários estímulos recebidos. No caso específico da leitura, segundo o conexionismo, quando “acessamos nosso léxico”, ou seja, a memória semântica, estamos de fato reinstanciando conexões interneuroniais a partir do estímulo visual obtido, a palavra impressa. No momento em que são reinstanciadas, essas conexões ativam, a um só tempo, diversas informações acerca desse estímulo (ortográfica, fonológica, semântica, pragmática,...) através de um processo denominado “memória auto-associativa” (IZQUIERDO, 2004; CIELO, 2004; CHIELE, 2004).

Na perspectiva de Christiansen e MacDonald (2002), as capacidades de processamento da linguagem emergem da experiência e dessa arquitetura de rede, e os autores afirmam que a performance no processamento da linguagem e a performance em atividades de memória de trabalho lingüística estão freqüentemente relacionadas uma com a outra.

A leitura compreensiva requer mais do que a decodificação ou o processamento isolado de palavras, ela também exige que o leitor processe o sentido das orações, estabeleça relações causais entre elas e faça inferências interagindo com o texto, construindo sentido e se apropriando do conteúdo do mesmo. O leitor de posse do seu conhecimento prévio e a partir das inferências que faz durante a leitura precisa integrar a informação do texto em unidades maiores de sentido para que consiga assim, aprender com a leitura.

A perspectiva conexionista da leitura busca integrar dados comportamentais, neurológicos e computacionais (SEIDENBERG e McCLELLAND, 1989). Por isso consideramo-la a mais adequada para emoldurar a presente pesquisa que, apesar de não produzir dados neurológicos e computacionais, ampara-se nos dados disponíveis na literatura.

O modelo cognitivo de leitura textual de Kintsch e van Dijk (1978) – posteriormente expandido e atualizado por vários pesquisadores, entre eles os próprios autores: Kintsch e van Dijk (1983) e Kintsch (1998) lança luz sobre o papel das habilidades de memória, linguagem e conhecimento de mundo, apresentadas na seção de aspectos cognitivos derivados do leitor, mas também ressalta a importância da construção de uma representação

textual e da necessidade de o leitor fazer inferências no processamento do texto em prol de uma leitura compreensiva geradora de aprendizagem. O modelo de Kintsch e van Dijk merece ser revisitado com um olhar conexionista por três motivos: a) por ser provavelmente o modelo que mais exerce influência nas pesquisas sobre leitura atualmente, uma vez que propõe um processo unificado que engloba a maioria das características da leitura descritas nos demais modelos; b) por ser aquele que melhor dá conta, a nosso ver, do processo de leitura e c) por que as simulações conexionistas amparam-se em dados comportamentais confiáveis e, em se tratando de leitura, o trabalho desses autores fornece esses dados.

Conforme Salles (2002), o Modelo de Compreensão de Textos proposto por Kintsch e van Dijk (1978) sugere que, devido às limitações na memória de curto prazo, o processamento de um texto é feito em ciclos, que correspondem, aproximadamente, a uma frase. No modelo, o texto é transformado em proposições (unidades de informação) que são combinadas pelo leitor. Desses ciclos, uma estrutura mais geral de significado, a macroestrutura, é extraída das proposições do texto original e mantida na memória episódica junto com os itens finais do ciclo. As frases apresentadas no texto são conhecidas como microestruturas. O ciclo seguinte altera as representações dos anteriores na memória episódica, fazendo com que ocorra a construção gradual de um texto base.

Porém, a noção de macroestrutura empregada por Kintsch e van Dijk (1978) para definir a descrição semântica do conteúdo global do discurso tem sua abordagem baseada no paradigma simbólico, conforme se pode verificar nas explicações dos processos de construção, armazenagem e recuperação dessa estrutura. A macroestrutura a partir do paradigma simbólico pode ser entendida como uma certa organização de significado que representa os aspectos essenciais de um texto, por meio de unidades denominadas macroproposições. A compreensão pode ser entendida como um processo que permite elaborar a macroestrutura do texto a partir de sua microestrutura. No processamento e retenção da informação textual, as macroproposições têm maior probabilidade de serem lembradas e incluídas em resumos (KINTSCH, 1998).

Segundo Kintsch (1998), a compreensão é a construção de uma estrutura. Compreender um texto significa formar uma estrutura mental que represente o sentido e a mensagem do texto. Nesse modelo, o leitor constrói durante a leitura uma representação do

conteúdo do texto (um modelo mental do texto) a partir de dados como o contexto, personagens, objetos, causas, intenções, entre outros elementos.

Na visão do conexionismo a macroestrutura, que também é entendida como necessária para a compreensão, não é tida como necessária ou imprescindível para a armazenagem do conteúdo textual. Esse paradigma, ao revisitar a noção de macroestrutura, não a vê como uma estrutura rígida e estanque, armazenada em um local específico da memória e recuperada posteriormente da mesma maneira com que foi gravada. Essa visão está baseada no fato de que o leitor pode recuperar a macroestrura de forma diferente da que foi gravada e que a cada vez que a recupera, pode recuperá-la diferentemente. Somado a isso está o fato de que o leitor não recupera a macroestrutura de uma vez só, mas sim faz uma retomada do assunto aos poucos (CASTRO, 2004).

Assim, no conexionismo a macroestrutura é formada, segundo Castro (2004), a partir de todos os dados textuais que ativaram as conexões interneuroniais mas não é formada para que tais dados sejam armazenados.

Segundo Chiele (2004), as pesquisas implementadas em uma perspectiva cognitiva reconhecem a existência de mecanismos de identificação de palavras de naturezas diferentes, que se modificam no decorrer do tempo. O mecanismo inicial de identificação de palavras através do processo de recodificação tende a evoluir em decorrência da automatização progressiva: leitores habilidosos caracterizam-se, entre outros aspectos, pela capacidade de acessar diretamente o significado das palavras lidas (FARR e CAREY, 1986).

Isso se dá porque, segundo o conexionismo, o conhecimento de cada palavra é codificado de maneira distribuída em engramas. A leitura de palavras envolve, essencialmente, o acesso a três tipos de código: ortográfico, fonológico e semântico. A construção da representação em cada um desses níveis influencia e é influenciada pelos demais.

Segundo a perspectiva conexionista, os três níveis de informação referentes a uma determinada palavra são sempre ativados durante a leitura em paralelo, mesmo os que não são diretamente requeridos para a execução de uma dada tarefa (SALLES, 2002).

O leitor pode manipular seu modelo mental durante a leitura, reordenando suas partes ou combinando-as para obter outras relações num processo mental de recorte e colagem, tudo isso para poder ter uma representação do conteúdo do texto ao final da leitura. Por exemplo, é freqüente que a ordem canônica dos acontecimentos de um texto seja alterada por razões estilísticas ou expressivas; a memória do sujeito, porém, tende a reordenar as partes do texto de acordo com as relações causais e temporais implícitas entre essas (KINTSCH e VAN DIJK, 1983). O sujeito faz isso construindo relações significativas entre as partes do texto (cadeias coesivas), através da compreensão da relação entre pronomes e referentes e outras formas de anáfora, bem como através do agrupamento de idéias positivas e negativas ao longo do texto.

Essa construção se dá diferentemente em cada indivíduo, não tendo como ser aplicada em série e de forma padronizada. Embora sejam as mesmas operações elementares que estabelecem as bases para a compreensão em leitura, sabe-se que cada indivíduo tem uma modalidade própria de construção e representação do conhecimento e é esse conhecimento representado durante a leitura que, segundo Poersch (1999), será utilizado para compreender o texto.

> É no cérebro que os dados trazidos pelo texto, juntamente com os dados previamente armazenados, começam a ser processados. Como conhecimento significa conexões sinápticas, se os dados de input encontram o caminho ou a conexão até os dados armazenados, esses dados armazenados são ativados. Nós dizemos que houve lembrança e sinapses já existentes são reforçadas. Se os dados de input não encontram um caminho previamente trilhado eles precisam ser integrados em algum dado já armazenado. Essa integração significa criação de uma nova conexão; em outras palavras, significa aprender. Então, esse conhecimento torna-se parte do conhecimento prévio e ambos são usados para processar o restante do texto[12].

Diferentes teorias da compreensão do texto, como as de Just e Carpenter (1987), têm visões diferentes acerca das características precisas dessa estrutura mental, mas todos convergem para a construção de tal estrutura de representação.

---

[12] No original: *It is in the brain that the data coming from the text, along with the data previously stored, start to be processed. As knowledge means synaptical connections, if input data find the path or the connection onto other stored data, these stored dada are activated: we say that* ***recall*** *occurs and the previous synapses are reinforced. If the input data do not find a previously trailed path, they have to be integrated into some already stored data. That integration consists of setting up a new connection; in other words, it means* ***to learn****. Then this knowledge becomes part of the prior knowledge and both are used to process the remaining part of the text* (p. 86-87).

Essa representação, a qual não se tem acesso direto, tem sido arena de muitos embates. Nesse espectro tem-se apresentado como um desafio o estabelecimento de níveis que permitam vislumbrar com maior nitidez as representações de sentido que os leitores constroem durante a compreensão.

Graesser e colaboradores (1997) apontam que o campo do processamento do discurso vem analisando muitos dos vários níveis de representação que são construídos no momento em que o sujeito lê ou ouve um discurso coerente. Psicólogos do discurso desenvolveram modelos que especificam como esses níveis são mentalmente representados e como eles são dinamicamente construídos durante a compreensão.

Vários níveis de representação do discurso foram identificados por pesquisas em lingüística textual e computacional, sociolingüística e estudos literários. Porém, alguns desses níveis não foram aceitos por psicólogos do discurso por serem, segundo Graesser e colaboradores (1997), aplicáveis a um conjunto muito limitado de contextos de discurso.

Kintsch e Van Dijk (1978) apresentam em seu modelo uma distinção de três níveis na representação do discurso: 1) a estrutura de superfície; 2) a base textual (contendo o conteúdo semântico local e global), e 3) o modelo de situação referencial.

Segundo os autores, o primeiro nível preserva o exato vocabulário e sintaxe da frase. Normalmente os leitores retêm apenas o código de superfície das frases mais recentes, a menos que esse código tenha repercussão importante no sentido.

A compreensão de um texto implica a criação de um texto base, ou de uma base textual (segundo nível), que organiza hierarquicamente todas as proposições[13] explícitas no texto, combinando conhecimentos sobre a linguagem e conhecimentos sobre o mundo (KINTSCH, 1998). A construção de tal base textual implica na elaboração de uma série de inferências que proporcionam a coerência semântica local entre as proposições (GRAESSER et al, 1997)

---

[13] Proposições são unidades abstratas de significado e implicam, no mínimo, a predicação de algo (uma propriedade, ação, relação, etc.) acerca de algo (um objeto, argumento, etc.) (KINTSCH, 1998, p. 37-38).

O terceiro nível é o modelo de situação, que consiste no conteúdo ou no micromundo sobre o qual o texto trata, ou seja, é a representação mental do texto. O modelo de situação de uma história refere-se às pessoas, cenário, ações e eventos no micromundo mental. Esse micromundo é construído inferencialmente por meio de interações entre o texto explícito e o conhecimento de mundo anterior.

Segundo Kintsch (1998), o foco de muitas pesquisas que vêm sendo feitas está na construção do modelo de situação (Graesser e Zwaan, 1995; Zwaan, Langston e Graesser, 1995; Trabasso e Suh, 1993; van den Broek, Risden, Fletcher, e Thurlow, 1996) e, conforme o autor, fica claro a partir de todas essas pesquisas que não há um único tipo de modelo de situação e nem um único processo para a construção de tal modelo.

Os modelos de situação, nos quais a situação descrita no texto é representada, são como os modelos mentais, ou seja, um construto localizado na memória episódica que contém a representação subjetiva do texto, incluindo o processo inferencial e experiências prévias.

O processo inferencial viabiliza a organização dos sentidos elaborados pelo indivíduo na sua relação com o texto. É a partir dele que o estabelecimento da relação entre as partes do texto e entre essas e o contexto torna-se possível, fazendo dele uma unidade aberta de sentido.

Segundo Ferreira e Dias (2004), além de favorecer a organização das relações de significado dentro do texto, o processo inferencial permite destacar a *malha* ou *teia* de significados que o leitor é capaz de estabelecer dentro do horizonte de possibilidades que é o texto. Essas relações não são aleatórias, mas se originam no encontro-confronto de dois mundos em situação de leitura: o do autor e o do leitor.

Nessa atividade, o conhecimento geral do sujeito-leitor exerce papel fundamental. É lançando mão desse conhecimento (lingüístico e de mundo) que o leitor poderá garantir uma compreensão para além dos elementos superficiais do texto, negociando com o autor os significados plausíveis e permitidos.

É por isso que é aqui destacada, em convergência com alguns autores da área da ciência do texto referidos na literatura, tais como Cain, Oakhill, Barnes e Bryant, 1998; Johnson e Smith, 1981; Marcuschi, 1985, 1989; Oakhill e Garnham, 1988; Oakhill, Cain e Yuill, 1997; Yuill e Oakhill, 1991, a atividade inferencial como sendo um 'fator essencial' no processo de compreensão em geral, podendo a mesma ser descrita como um ato inteligente que envolve raciocínio lógico e criativo, e que é levado a efeito através da junção de informações novas e antigas, possibilitando o surgimento de novas intuições ou conclusões (SPERBER, 1996).

Assim, o paradigma conexionista, revisitando o modelo de Kintsch e van Dijk (1978), ao propor essa mudança de enfoque, no qual a compreensão em leitura deriva da construção de uma representação mental, acaba trazendo algumas importantes contribuições para se pensar as formas de avaliação da compreensão e, conseqüentemente, para o ensino da leitura proficiente, uma vez que os testes, especialmente no caso de L2, servem como parâmetro para o ensino. A avaliação da compreensão leitora será abordada na próxima seção.

1.3 Avaliação da compreensão leitora

> O processo de avaliar [...] é mais efetivo e alcança melhores resultados quando se baseia em princípios operacionais adequados. Esses princípios, ou guias, proporcionam certa direção ao processo e servem como critérios normativos para avaliar a efetividade dos procedimentos e das práticas que se utilizam para ele[14]. (Fermín, 1971)

Diante de tudo que foi escrito até aqui acerca da compreensão, é indiscutível que a avaliação da mesma não se constitui em uma tarefa simples, pois como processo mental que é, seu acesso é sempre indireto. Além dessa dificuldade, a profusão de elementos e aspectos cognitivos envolvidos também colabora para tornar a ação avaliativa mais difícil. Porém, como dito na seção anterior, o conhecimento acerca desses aspectos colabora com a tarefa, servindo de base para que se decida "o quê" e "como" medir, além de servir de indicador das capacidades cognitivas que devem ser consideradas no momento de avaliar a compreensão.

[14] No original: *El proceso de evaluar [...] es más efectivo y alcanza mejores resultados cuando se basa en adecuados principios operacionales. Estos principios, o ideas-guías, proporcionan cierta dirección al proceso y sirven como criterios normativos para apreciar la efectividad de los procedimientos y da las prácticas que se utilizan para ello* (p.23).

A avaliação da compreensão leitora revela inegavelmente um olhar sobre a compreensão. A concepção de compreensão enquanto uma lista de habilidades, por exemplo, resulta em uma avaliação que mede tais habilidades isoladas. Esse tipo de avaliação, bastante tradicional em nosso meio, vem recebendo críticas ao longo do tempo.

Segundo alguns psicolingüistas, como Cooper e Petroski (1976), citados por Alliende e Condemarín (2005, p. 132), as habilidades contidas nas listas consistem em categorias muito amplas, gerais e simplistas, válidas para qualquer processo cognitivo, mas não relacionadas com a linguagem, nem específicas do ato de ler. Outra crítica apresentada ainda em Alliende e Condemarín (2005), referente às listas de sub-habilidades, é que as mesmas estariam elaboradas a partir de um modelo psicométrico e consistiriam assim em simples rótulos empregados por diferentes autores de testes para o tipo de perguntas elaboradas por eles. No entanto, as críticas feitas a essa concepção não devem deslegitimar a importância de se avaliar e ensinar habilidades de compreensão da linguagem escrita.

Como o propósito nesta pesquisa é observar a avaliação da compreensão especificamente em língua estrangeira, na próxima sub-seção serão apresentadas as diferentes visões nessa arena específica da L2.

1.3.1 Avaliação da compreensão leitora em L2

Conforme Koda (2005), a tendência atual na avaliação da compreensão em L2 se focaliza mais na proficiência da língua do que na compreensão leitora. A autora afirma que apesar de a fluência na língua ser desejável, um cuidado especial deve ser tomado ao avaliar, dando um destaque a isso, pois a eficiência no processamento de segunda língua ou língua estrangeira nem sempre retrata a habilidade de *leitura* na língua alvo.

Segundo a autora, a avaliação da compreensão leitora em L2 com foco na proficiência está muito mais embasada em aspectos da legibilidade física e lingüística do texto apresentados na seção 1.2.3.2, julgando–os como sendo determinantes mais precisos da dificuldade na compreensão do que elementos não-lingüísticos.

Koda (2005) aponta ainda que as diretrizes sobre proficiência desenvolvidas pelo Conselho Americano para o Ensino de Línguas Estrangeiras (ACTFL) definem a proficiência

em leitura com base nos tipos de texto e numa listagem de habilidades associadas com cada tipo, como pode ser visto no Quadro 3 e, ao fazerem isso, conferem um papel proeminente à complexidade lingüística implícita no tipo textual.

Nível	**Tipo de texto**	**Amostras de textos**	**Habilidades de leitura**
0	De enumeração	Números, nomes, placas de rua, denominação de dinheiro, designações de escritório/compras, endereços	Reconhecimento de elementos memorizados
1	Orientada	Formulários de viagem e registro, horário de trens e aviões, guias de programas de tv e rádio, menus, memorandos, manchetes de jornal, quadros de conteúdo, mensagens	Folhear (*skim*) e examinar (*scan*)
2	Instrutivo	Comerciais e rótulos, relato de jornal, instruções e direções, relatos reais, solicitações em formulários, convites, parágrafos introdutórios e conclusivos	Decodificação e classificação
3	Avaliativo	Editoriais, análises, apologias, textos literários, biografia com interpretação crítica.	Inferência, adivinhação, levantamento de hipótese e interpretação
4	Projetivo	Críticas de arte ou performance de teatro, textos literários, discursos filosóficos, material técnico, argumentação.	Análise, verificação, amplificação, levantamento de hipótese

Quadro 3 - Níveis de leitura associados com tipos de texto e habilidades de processamento
Fonte: adaptação KODA (2005, p. 248).

Conforme Koda (2005), vários pesquisadores, como Allen, Bernhardt, Berry, e Demel (1988); Lee e Musumeci (1988) se posicionam contrariamente a tais diretrizes, argumentando que a performance em leitura é influenciada por outros fatores, não somente variáveis lingüísticas, e que a hierarquia do tipo textual corresponde a algumas, mas não a todas, as dimensões da proficiência em leitura de língua estrangeira.

Sob o prisma do conexionismo, como já dito, a avaliação da compreensão leitora tem seu foco alterado: não se trata mais de avaliar o processo de compreensão em leitura a partir do estudo de estratégias (e habilidades) isoladas e generalizadas como se a simples soma das partes pudesse refletir o todo, uma vez que, conforme Morin (1991, p. 18-19), “o todo é mais do que a soma das partes”. Ao invés disso, a avaliação nesse paradigma prioriza a investigação das múltiplas relações entre as diversas operações e estratégias colocadas em jogo durante a leitura de um texto para a formação da representação mental do conteúdo.

Fica, assim, evidente que a questão toda está na *lente* utilizada no momento de olhar para a compreensão e, conseqüentemente, para a sua avaliação: se pensarmos na compreensão como o processamento das várias habilidades pertinentes e se esse processamento for elaborado a partir de um referencial teórico produtivo, estar-se-á contribuindo para a construção de instrumentos de avaliação consistentes que informem a respeito da autêntica habilidade de compreensão do candidato, para os devidos fins aos quais a avaliação se propõe.

Nas sub-seções seguintes serão apresentados os propósitos da avaliação, alguns instrumentos tradicionalmente utilizados para medir a compreensão leitora, as etapas de validação desses instrumentos e os critérios utilizados para validar os mesmos.

1.3.2 Propósitos da avaliação

Segundo Koda (2005), os propósitos de avaliação e os usos que se pretende com os resultados dessa prática deveriam guiar o *design* das tarefas presentes nesses testes. A autora aponta três funções avaliativas como as mais comuns: administrativa, de diagnóstico e classificatória.

A primeira diz respeito a decisões políticas e administrativas e cumpre função de designações. A segunda tem como objetivo central identificar fontes de dificuldade na leitura. A última, que tem por propósito determinar habilidades de leitura individuais, cria base legítima para comparações entre os *testados* e fornece dados que podem ser utilizados para diferenciar candidatos.

Segundo Alliende & Condemarín (2005), também é possível avaliar a compreensão para adaptar o ensino aos níveis de habilidade de leitura e às necessidades individuais ou de grupo. Com esse propósito, a avaliação estaria servindo para problematização do ensino e da educação.

1.3.3 Instrumentos utilizados para medir a compreensão em leitura

> Sofremos tristemente esses contínuos exames. “Estropeiam” nossos jovens, que não têm tempo nem ocasião para serem preguiçosos, preguiçosos no bom sentido.... Porque eu acredito na preguiça inteligente, que dá ao homem tempo de ler para si. Não se pode chamar estudo ao que um homem deve ler página por página, especificamente para prestar um exame[15]. (FERMÍN, 1971)

Como a compreensão em leitura é uma atividade complexa e com muitos elementos envolvidos, é coerente se pensar que um só instrumento de avaliação não conseguiria abarcar todos os aspectos. Assim, vários instrumentos podem ser utilizados para avaliar a compreensão em leitura, dependendo dos critérios e propósitos de avaliação vistos anteriormente. As ações avaliativas, segundo Alliende e Condemarín (2005), podem ser feitas mediante procedimentos qualitativos ou informais e mediante medidas formais, como as referentes a critérios, a avaliações taxonômicas ou a testes padronizados.

1.3.3.1 Técnicas de avaliação informal

Segundo Koda (2005), várias técnicas de avaliação informal são usadas em sala de aula, como **a análise oral dos enganos** (o*ral miscue analysis*) e **a sondagem através da observação** (o*bservation survey*). A vantagem de utilizar essas técnicas, conforme a autora, está na riqueza dos dados derivados de uma ampla visão dos comportamentos de leitura fornecidos durante seu desenvolvimento.

A análise oral dos enganos (o*ral miscue analysis*) consiste em analisar os erros cometidos durante a leitura oral de um texto. No entanto, a autora aponta como um fator limitante dessa abordagem o fato de ela enfocar apenas um aspecto do comportamento de leitura – os erros na leitura oral – para inferir a habilidade de compreensão na leitura silenciosa.

A sondagem através da observação (o*bservation survey*) tem como objetivo obter um entendimento mais profundo do que acontece com os indivíduos em conjunto com o contexto educacional. Essa abordagem consiste em um relatório a ser preenchido mediante a observação individual de sujeitos, realizando uma série de tarefas que medem as várias

[15] No original: *Sufrimos tristemente estos continuos exámenes. Estropean a nuestros jóvenes, que no tienen tiempo ni ocasión para ser perezosos; perezosos en el buen sentido… Porque yo creo en la pereza inteligente, que da al hombre tiempo de leer para sí. No se puede llamar estudio a lo que un hombre debe leer, página por página, exactamente para sufrir un examen* (p. 100).

capacidades entendidas como requisitos para o desenvolvimento da leitura. As limitações desse instrumento estão na demanda de tempo e trabalho.

1.3.3.2 Técnicas de avaliação formal

> Este tipo de avaliação indica se o aluno que está sendo avaliado maneja ou não uma habilidade específica de acordo com um padrão estabelecido. Os testes relacionados a critérios, geralmente, associam-se a objetivos procedimentais e podem ser úteis para individualizar o ensino da leitura (HARRIS e SIPAY, 1979).

As técnicas de avaliação formal consistem nos testes padronizados, relacionados a critérios, utilizados para avaliar a compreensão. Conforme Alliende e Condemarín (2005), mediante o enfoque desses testes são avaliadas as realizações e dificuldades dos alunos em relação a determinados objetivos do processo de ensino-aprendizagem, e não em relação ao rendimento obtido pelos estudantes.

Ainda segundo os autores, a principal vantagem das medições relacionadas a critérios é que mostram diretamente o rendimento dos alunos tendo como referência objetivos específicos de ensino. Porém, suas limitações dizem respeito ao fato de não haver critério estável sobre a natureza do domínio de aprendizagem e sobre como medir esse domínio, uma vez que se trata de um processo mental, como já mencionado anteriormente.

Outro problema apresentado pelos autores em relação a tais medições é determinar se o critério de 80 a 90% de objetivos alcançados, geralmente empregado, indica na realidade domínio de uma determinada aprendizagem, ou se deveriam existir outros critérios, como determinar a importância de uma competência para conseguir transferência da aprendizagem para outras.

Como exemplo de medida formal referente a critérios é apresentado o teste Cloze. Criado por Wilson L. Taylor e apresentado pela primeira vez em 1953, esse instrumento gera muita polêmica entre os pesquisadores quanto ao quê de fato ele mede e quais são suas possíveis aplicações (GABRIEL, 2005, p. 183).

O suporte psicológico do teste Cloze vem da teoria de *Gestalt,* que se refere à habilidade humana em completar os elementos omitidos de uma totalidade estruturada. O teste consiste em solicitar ao candidato que insira certas palavras suprimidas do texto. As

omissões são feitas em intervalos regulares de cinco em cinco palavras, por exemplo. A palavra omitida deve ser substituída por uma linha de tamanho constante. Normalmente as palavras omitidas não devem ser nomes próprios. Na primeira e última oração do texto não se omitem palavras.

O projeto básico do teste Cloze pode adotar diversas modificações. Conforme Koda (2005), o procedimento parece ser mais sensível a formas de superfície lingüística do que ao sentido subjacente do texto, pois dependendo de qual palavra for apagada primeiro, palavras totalmente diferentes vão requerer recolocação. Um estudo apresentado em Alderson (2000) verificou que o desempenho de alguns indivíduos modificou consideravelmente em versões alternativas do teste criadas a partir do mesmo texto.

Alguns pesquisadores divergem quanto à aceitação ou não de sinônimos ao invés da resposta exata no preenchimento das lacunas. Porém, a maioria dos autores, conforme Alliende & Condemarín (2005, p. 131), “aceita qualquer vocábulo da mesma família semântica que complete adequadamente o sentido do texto”.

Quanto aos índices de escore do Cloze, segundo Bormuth (1968), as respostas corretas entre 44 e 57% indicam um nível instrucional médio; as pontuações abaixo de 44% indicam não compreensão do texto e acima de 57% revelam nível de compreensão proficiente.

Outro procedimento de avaliação são as taxonomias textuais. Para abarcar a totalidade dos aspectos dos conteúdos elaboraram-se algumas taxonomias textuais. Conforme Alliende e Condemarín (2005), as taxonomias tentam indicar as categorias em que se podem incluir os componentes de um texto. Os autores ressaltam que, como os textos tendem à variedade, nunca conseguem esgotar as possibilidades em um só texto. Por isso, as taxonomias devem ser consideradas como sugestões que ajudam a encontrar as verdadeiras categorias de análise dos referentes de um texto.

Um bom exemplo é constituído pela taxonomia de Barret (1968), apresentada na seção 1.3 desse capítulo. Originalmente trata-se de uma taxonomia construída especialmente para textos narrativos, da qual se podem tirar idéias para outros tipos de textos. A seguir é apresentado um quadro com uma síntese e adaptação dos principais pontos da taxonomia.

<table>
<tr><td colspan="2">COMPREENSÃO LITERAL
A compreensão literal refere-se à recuperação da informação explicitamente apresentada no texto. Um exemplo de tarefa simples na compreensão literal pode ser o reconhecimento ou lembrança de um único fato ou incidente. Um exemplo de tarefa mais complexa pode ser o reconhecimento ou lembrança de uma série de fatos ou a seqüência de incidentes. Propósitos e questões nesse nível podem ter as seguintes características:</td></tr>
<tr><td>Reconhecimento:</td><td>consiste na localização e identificação de elementos do texto.</td></tr>
<tr><td>Reconhecimento de detalhes:</td><td>requer localizar e identificar fatos como nomes de personagens, incidentes, tempo e local do conto.</td></tr>
<tr><td>Reconhecimento de idéias principais:</td><td>requer localizar e identificar uma oração explícita no texto, que seja a idéia principal de um parágrafo ou de um trecho mais extenso da seleção.</td></tr>
<tr><td>Reconhecimento de seqüências:</td><td>requer localizar e identificar a ordem de incidentes ou ações explicitamente apresentadas no trecho selecionado.</td></tr>
<tr><td>Reconhecimento de comparação:</td><td>requer localizar ou identificar similaridades e diferenças em personagens, épocas e lugares explicitamente apresentados no trecho selecionado.</td></tr>
<tr><td>Reconhecimento das relações de causa e efeito:</td><td>requer localizar ou identificar as razões que, estabelecidas com clareza, determinam um efeito.</td></tr>
<tr><td>Reconhecimento de traço de personagens:</td><td>requer localizar ou identificar declarações explícitas sobre um personagem que ajude a destacar de que tipo de pessoa se trata.</td></tr>
<tr><td>Lembrança:</td><td>requer que o leitor reproduza de memória idéias e informações explicitamente apresentadas no trecho.</td></tr>
<tr><td>Lembrança de detalhes:</td><td>requer reproduzir de memória fatos como nomes de personagens, tempo e local do conto, fatos minuciosos.</td></tr>
<tr><td>Lembrança de idéias principais:</td><td>requer lembrar-se da idéia principal de um texto, principalmente quando essa se encontra expressamente estabelecida. Também se pode referir às idéias principais de alguns parágrafos.</td></tr>
<tr><td>Lembrança de seqüências:</td><td>consiste em lembrar-se da ordem dos incidentes ou das ações apresentados com clareza no texto.</td></tr>
<tr><td>Lembrança de comparações:</td><td>requer lembrar-se de memória as semelhanças e diferenças nos personagens, época e lugares explicitamente apresentados no texto.</td></tr>
<tr><td>Lembrança de relações de causa e efeito:</td><td>requer lembrar-se das razões que determinam um efeito.</td></tr>
<tr><td>Lembrança de traços de personagens:</td><td>requer lembrar-se da caracterização explícita que foi feita dos personagens que aparecem no texto: nomes de personagens, tempo e local do conto.</td></tr>
<tr><td colspan="2">REORGANIZAÇÃO
A reorganização consiste em analisar, sintetizar e/ou dar uma nova organização às idéias, às informações explicitamente apresentadas no trecho selecionado.</td></tr>
<tr><td>Classificação:</td><td>consiste em situar em categorias as pessoas, os objetos, os lugares, e as ações mencionadas no texto, ou em exercer a atividade classificatória sobre qualquer elemento do texto.</td></tr>
<tr><td>Esboço:</td><td>consiste em reproduzir o texto de forma esquemática. Pode-se executar utilizando orações ou mediante representações ou disposições gráficas.</td></tr>
<tr><td>Resumo:</td><td>consiste em uma condensação do texto mediante orações que reproduzem os fatos ou elementos principais.</td></tr>
<tr><td>Síntese:</td><td>consiste em reunir diversas idéias, fatos ou outros elementos do texto em formulações mais abrangentes.</td></tr>
<tr><td colspan="2">COMPREENSÃO INFERENCIAL
A compreensão inferencial exige que o estudante use as idéias e as informações explicitamente apresentadas no trecho, a sua intuição e a sua experiência pessoal como base para conjecturas e hipóteses. As inferências podem ser de natureza convergente ou divergente, e o leitor pode ou não ser exigido a verbalizar a base racional de suas inferências. Em geral, a compreensão inferencial é estimulada mediante a leitura, e as perguntas que demandam pensamentos e imaginação que vão além da página impressa</td></tr>
<tr><td>Inferência de detalhes:</td><td>exige conjeturar sobre os detalhes adicionais que o autor poderia ter incluído no texto para torná-lo mais informativo, interessante ou atrativo.</td></tr>
<tr><td>Inferência de idéias</td><td>exige incluir a idéia principal, o significado geral, o tema ou ensino moral que</td></tr>
</table>

principais:	não estão expressamente apresentados no texto.
Inferência de seqüências:	consiste em determinar a ordem das ações, se a sua seqüência não é estabelecida com clareza no texto. Também pode consistir em determinar as ações que precederam ou se seguiram às que são observadas no texto.
Inferência de comparações:	requer inferências acerca de semelhanças e diferenças nos personagens, épocas e lugares.
Inferência de causa e efeito:	exige apresentar hipóteses sobre as motivações dos personagens e de suas interações com o tempo e o lugar. Também implica conjeturar sobre as causas que atuaram a partir de indícios explícitos apresentados.
Inferência de traços dos personagens:	consiste em determinar características dos personagens que não são dadas explicitamente no texto.
Predição de conseqüências:	requer que o leitor leia uma parte inicial do texto e com base nessa leitura ele é solicitado a conjeturar sobre o que virá no decorrer do texto.
Interpretar linguagem figurada:	requer que o leitor infira acerca do sentido a partir da linguagem figurada.
LEITURA CRÍTICA/ AVALIAÇÃO A leitura crítica exige que o leitor emita um julgamento de valor, comparando as idéias apresentadas no texto com critérios externos dados por ele, leitor, por outras autoridades ou por outros meios escritos, ou com um critério interno dado pela experiência, pelos conhecimentos ou pelos valores do leitor.	
Julgamento de realidade ou fantasia:	exigem que o leitor seja capaz de distinguir entre o real do texto e o que pertence à fantasia do autor.
Julgamento de fato ou opinião:	requer que o leitor analise e avalie a escrita do autor com base no conhecimento que ele tem do assunto bem como analise e avalie a intenção do autor.
Julgamento de adequação e validade:	requer que o leitor compare fontes escritas de informação observando concordância e discordância e integralidade.
Julgamento de apropriabilidade:	requer que o leitor faça um julgamento sobre a relativa adequação de diferentes partes do texto para responder a questão.
Julgamento de valor, desejo e aceitação:	requer que o leitor emita julgamentos baseado em seu código moral ou seu sistema de valores.
Apreciação:	implica todas as considerações prévias, pois lida com o impacto psicológico e estético que o texto produziu no leitor. Inclui o conhecimento e a resposta emocional às técnicas literárias, ao estilo e às estruturas.
Resposta emocional ao conteúdo:	requer que o leitor verbalize sobre seus sentimentos em relação ao texto em termos de interesse, empolgação, aborrecimento, medo, ódio, diversão, etc. Está relacionada com o impacto emocional da leitura no leitor.
Identificação com personagens ou incidentes:	As perguntas dessa natureza elicitam respostas que demonstrem sensibilidade, simpatia e empatia do leitor em relação aos personagens e acontecimentos relatados no texto.
Reações ao uso da linguagem do autor:	requer que o leitor responda sobre a habilidade do autor em termos das dimensões semânticas do texto, ou seja, conotações e denotações.
Imagem:	requer que o leitor verbalize seus sentimentos em relação à habilidade do autor em pintar palavras imagéticas que façam o leitor visualizar, escutar ou sentir o cheiro e o gosto.

Quadro 4 – Dimensões cognitivas e afetivas da compreensão leitora da Taxonomia de Barret (adaptação CLYMER, 1968, p. 19-23; ALLIENDE e CONDEMARÍN, 2005, p. 132-134)

Outra técnica formal muito utilizada para avaliação da compreensão leitora são as questões de múltipla-escolha. As vantagens estão na praticidade e simplicidade do seu procedimento de contagem dos resultados. A natureza mecânica dessa técnica não apenas garante a objetividade, mas também permite contagem mediada por equipamentos. Por outro

lado, a técnica apresenta desvantagens. É necessário cuidado na interpretação dos resultados obtidos com essa técnica, pois variações na performance poderiam ser mais atribuídas a habilidades de tarefas específicas do que à habilidade de leitura.

Um exemplo de teste que se caracteriza por oferecer em sua maioria questões de múltipla escolha medindo diferentes habilidades e níveis de compreensão é o TOEFL (*Test of English as a Foreign Language*), que vem a ser, entre os testes padronizados, um dos mais reconhecidos internacionalmente. Conforme informações no *site* [www.ets.org/toefl] desenvolvido pelo instituto responsável por sua elaboração desde 1963, o *Educational Testing Service* (ETS), o TOEFL "mede a habilidade de falantes não nativos de inglês em usarem e entenderem a língua inglesa da forma em que é falada, escrita e ouvida em ambientes de faculdades e universidades[16]". Esse teste será descrito aqui com maior riqueza de detalhes por ser sua seção de leitura objeto de análise da presente pesquisa. Ao longo do texto serão apresentadas algumas telas da versão CBT utilizada na pesquisa a título de ilustração.

O TOEFL é utilizado para avaliar a proficiência em inglês de candidatos cuja língua materna não é o inglês e os resultados são usados por mais de 2.400 faculdades e universidades nos Estados Unidos e Canadá para determinar se o nível de proficiência de um candidato é aceitável para a instituição em questão. O teste também é empregado por instituições em outros países em que o inglês é o meio de instrução. O teste é aplicado pelo ETS uma vez por mês em mais de 6.000 instituições e agências em 110 países para selecionar candidatos com as habilidades necessárias para serem bem sucedidos (ETS, 2007)[17]. O construto de leitura que é medido no teste de leitura do TOEFL não é exposto explicitamente na literatura do ETS (PEIRCE, 1992).

Cada teste aplicado é utilizado uma única vez e a equipe de desenvolvimento de testes no ETS produz um novo teste TOEFL para cada administração mensal. O Conselho de Política do TOEFL é composto por um comitê de examinadores, um comitê de pesquisa e um

---

[16] No original: *The Test of English as a Foreign Language™ (TOEFL®) measures the ability of nonnative speakers of English to use and understand English as it is spoken, written, and heard in college and university settings.*

[17] As informações aqui apresentadas estão baseadas em uma série de materiais preparatórios para o TOEFL como livros publicado pela editora Kaplan, um "manual" da série preparatória de testes produzido pela SBS, bem como da própria página do instituto responsável pela elaboração do teste, o ETS, que além de informações disponibiliza alguns testes para prática.

comitê de serviços. Cada um é responsável por diferentes áreas da atividade do programa (PEIRCE, 1992).

O teste é oferecido em diferentes formatos, isto é, em suportes diferentes: TOEFL *paper-based* (TOEFL PBT) é o teste feito no suporte tradicional (papel); TOEFL *computer-based* (TOEFL CBT) é o teste feito no computador; e, por último, o TOEFL *internet-based* (TOEFL IBT) é o teste feito através da *internet*. Essa última modalidade do teste ainda não se encontra disponível em todos os países. O TOEFL *paper-based* confere um escore que varia de 310 a 670, o TOEFL *computer-based* confere um escore de 0 a 300, enquanto o novo TOEFL *internet-based* confere um escore total numa escala de 0 a 120. Todos têm validade de apenas 2 anos.

O TOEFL *computer-based* é a modalidade do teste mediado pelo computador que conta com um grande banco de questões que foram classificadas quanto ao tipo, ao nível de dificuldade e à área de conteúdo. Em algumas das seções, o banco de dados seleciona o nível de dificuldade das questões de acordo com o rendimento do candidato, o chamado (Computer-Adaptive Testing, CAT). O computador inicia o teste com uma pergunta de nível médio de dificuldade, então, dependendo da resposta dada pelo candidato, o programa adapta as próximas questões, buscando no banco de dados outra questão que leve em consideração a performance anterior, ou seja, se o candidato acertou a anterior, a próxima questão será mais difícil, ou se errou virá uma mais fácil. A seção de compreensão em leitura não é adaptada pelo computador. Essa modalidade mediada pelo computador foi a escolhida na presente pesquisa

No desenvolvimento de provas com itens de múltipla-escolha, como no TOEFL, são utilizados alguns termos psicométricos que formam uma linguagem padrão, como visto no seguinte exemplo:

> A resposta em uma questão de múltipla-escolha é referida como
> (A) um item
> (B) um distrativo
> (C) uma opção
> (D) uma chave

As alternativas A, B, C e D são as *opções*. A resposta para a questão, nesse caso a letra D, é a *chave*. As respostas incorretas, nesse caso as letras A, B, e C são os *distrativos*. A questão (que no exemplo é *A resposta em uma questão de múltipla-escolha é referida como*) a *base*. A base mais as opções é chamada de *item*[18] (PEIRCE, 1992). Na tela a seguir são apresentados os termos psicométricos referidos.

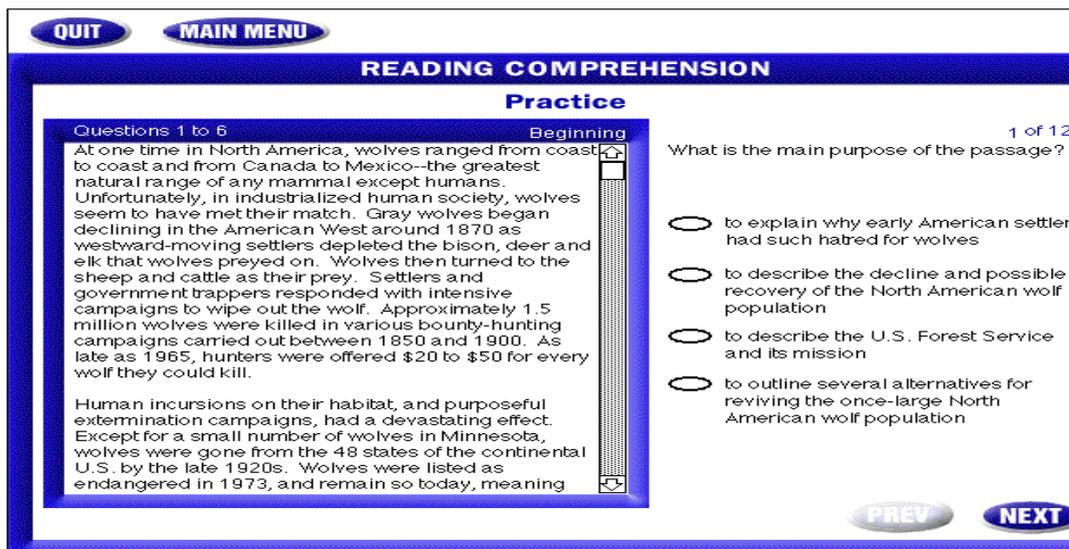



Figura 1 – Tela TOEFL ilustrando a linguagem psicométrica dos testes de múltipla escolha

A seção de compreensão leitora no TOEFL é dividida em cinco textos seguidos de seis itens cada um. Conforme informação no próprio material preparatório, as questões são classificadas da seguinte forma:

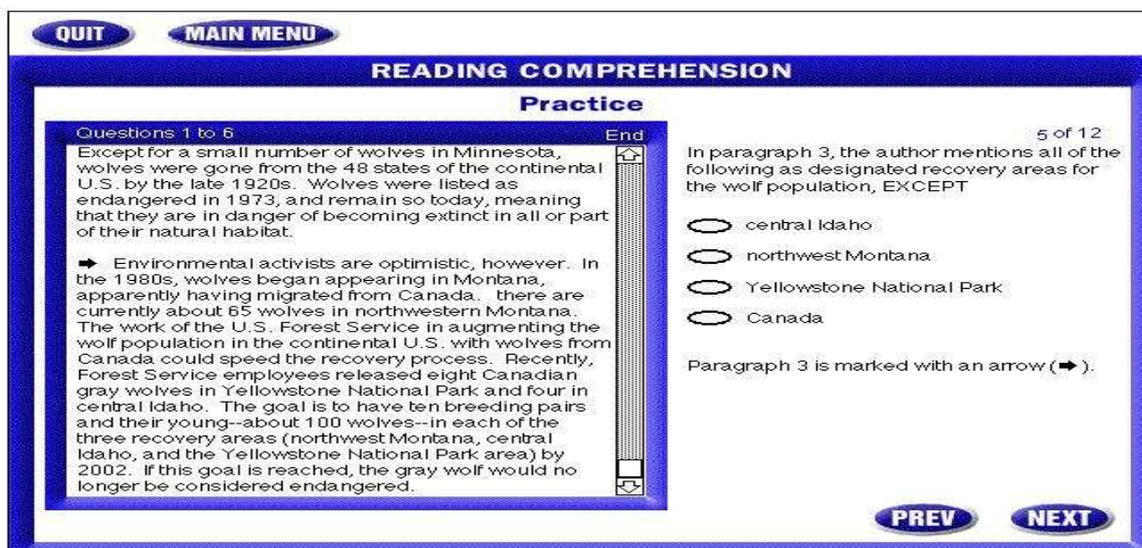
ado

ctoré
em

Figura 2 – Tela TOEFL questão 5

Tipo 2) Clicar na referência: nesse tipo de questão é dada uma palavra do vocabulário ou um pronome que aparece grifado no texto e o candidato é solicitado a clicar dentro do texto na palavra à qual aquela destacada se refere.

uma

e ser

Figura 4 – Tela TOEFL questão 3

Tipo 4) Inserir frase/clicar no *: nesse tipo de questão, o candidato tem que clicar em um

Figura 5 – Tela TOEFL questão 6

No TOEFL, os textos são chamados de passagens, uma vez que são apresentados sem o título e respectiva fonte e têm, em alguns casos, algumas adaptações feitas pelo elaborador do teste a fim de gerenciar o tempo do candidato.

Essas alterações, segundo Peirce (1992), que foi membro do departamento de desenvolvimento das provas do TOEFL no ETS de 1984 a 1987, ocorrem da seguinte forma: o "conteúdo que não era usado para propósitos de testagem era deletado do texto"[19]. Cabe aqui um contraponto: se a prova visa, conforme informação do *Bulletin of Information for TOEFL/CBT*, 2006 – 2007 (ETS, 2007), a utilização de material autêntico, semelhante aos textos acadêmicos utilizados em faculdades e universidades norte americanas, exatamente para poder verificar a habilidade do candidato em lidar com esses textos na vida real, não nos parece adequado alterar o texto retirando desse aspectos que o tornariam autêntico.

Além dessa quebra de autenticidade, outros aspectos merecem destaque. Testes padronizados de itens objetivos como o TOEFL apresentam a vantagem de ter uma correção rápida e prática, porém também apresentam desvantagens. Algumas de suas limitações, segundo Alliende e Condemarín (2005), dizem respeito ao fato de que, exatamente por serem padronizados em termos de sua administração e pontuação, podem ser inapropriados para a aplicação em certos sujeitos; o vocabulário, o tipo de perguntas e a carga conceitual costumam torná-los muito fáceis ou muito difíceis. Para Tumolo (2005), alguns itens desses testes padronizados, os chamados 'itens não-defensáveis', têm como conseqüência favorecer alguns grupos em detrimento de outros, já que conhecimento específico é pressuposto para algumas de suas respostas.

Outra limitação apresentada pelos testes padronizados do tipo múltipla-escolha, conforme Farr e Carey (1986), diz respeito à não dependência das questões em relação às passagens do texto. Isso se deve ao fato de que as respostas podem, muitas vezes, ser adivinhadas, tanto por mero acaso, o popularmente chamado "chute", quanto por indução da pergunta ou ainda eliminação de alternativas. Essa adivinhação pode acabar alterando os resultados, gerando com isso um escore não condizente com a habilidade efetiva de compreensão leitora do sujeito.

Essa característica presente em alguns testes acaba por ignorar um dos critérios centrais da compreensão leitora mencionado por Brown, citado por Farr & Carey (1986), que é a "qualidade de leitura"[20]. Em casos de testes cronometrados, como no caso do *TOEFL CBT*, o afastamento da qualidade de leitura pode vir a ser maior e acabar por incentivar direta

[19] No original: *content* [...] *that was not used for testing purposes was deleted from the text*. (p. 670).
[20] No original: *A key criterion mentioned by Brown (1914) [...] was "quality of reading". (p.30)*

ou indiretamente o candidato a adivinhar a resposta. Para os autores, a adivinhação é um fator que afeta tanto a validade quanto a confiabilidade do resultado do teste. É impossível determinar qual percentagem do teste pode ser atribuída às adivinhações feitas pelo candidato e qual percentagem pode ser garantida como uma resposta elaborada pelo candidato. Para contornar tal situação, os autores sugerem penalizar o candidato por adivinhar uma resposta e gerar instruções nos testes que incentivem os candidatos a evitar a técnica da adivinhação.

Porém essa sugestão vai de encontro àquelas vistas em alguns materiais didáticos preparatórios para o teste *TOEFL*, como a que pode ser vista a seguir:

> Sobretudo, **trabalhe rapidamente** de modo que você terá chance de responder todas as questões corretamente. Porém, se o tempo estiver acabando não há penalidade em adivinhar na seção de leitura, então certifique-se de que você respondeu todas as questões mesmo que você tenha que adivinhar ao acaso no final[21] (HILKE e WADDEN,2001).

Ainda segundo Farr e Carey (1986), outro grande problema dos testes com formato múltipla-escolha é a insistência em uma única resposta correta. Para os autores, que comungam do olhar dos Modelos Mentais de Kintsch e van Dijk (1978), a compreensão leitora é um processo construtivo; logo, o significado é tão dependente do leitor na criação de um modelo situacional quanto o é do texto.

Muitos dos problemas existentes nos testes de múltipla-escolha são de ordem técnica e estão relacionados a noções de probabilidade, estatística e produção de itens. Para Farr e Carey (1986), “a maioria dos testes de compreensão em leitura reflete teorias grosseiramente inadequadas acerca da cognição, linguagem e aprendizagem”[22]. Segundo os autores, embora não haja uma definição consensual sobre compreensão, são poucos os pesquisadores que conceberiam uma definição de compreensão baseada na palavra, como função da memória de curto prazo ou unidimensional.

Outra característica comum em testes de múltipla-escolha apontada por diversos autores, entre eles Hill e Parry (1992), Peirce (1992) e Farr e Carey (1986), é o fato de que, à medida que esses testes procuram não favorecer candidatos de áreas específicas, eles

---

[21] No original: *Above all, **work quickly** so that you give yourself a chance to answer all the questions accurately. If you do run out of time, however, there is no penalty for guessing in the Reading section, so make sure you answer every question even if you must guess randomly at the end (p. 11)* .Grifo do original.

[22] No original: *Most multiple-choice reading comprehension tests reflect grossly inadequate theories of cognition, language, and learning* (p. 35).

desestimulam os candidatos a utilizarem seus conhecimentos prévios. No caso do TOEFL, conforme o boletim de informações (ETS, 2007), é solicitado aos candidatos que respondam às questões de múltipla-escolha com base no que está “declarado” ou “implícito” em cada um dos textos (p. 21), sugerindo uma não necessidade de utilização de qualquer conhecimento prévio para responder a questão. Os candidatos devem escolher a que eles consideram ser a melhor opção entre as quatro apresentadas em cada item.

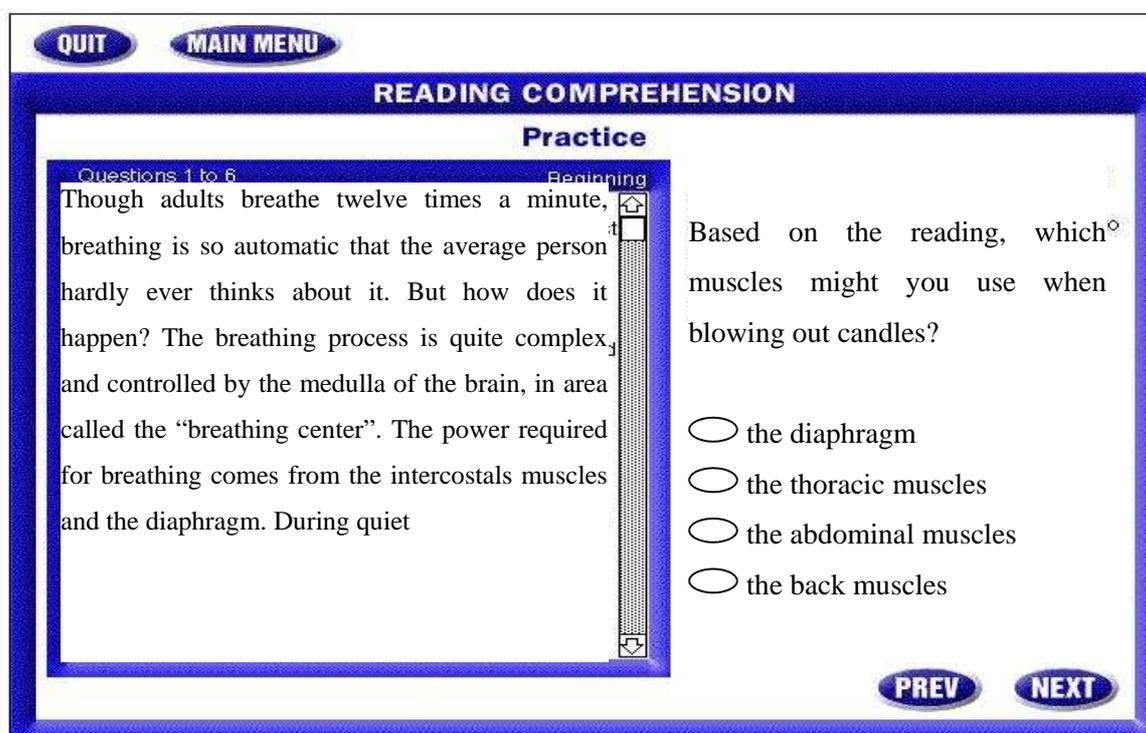
Tal postura pode acabar incidindo no comportamento de leitura dos candidatos, que muitas vezes suprimem sua bagagem de conhecimento pessoal por medo de marcarem uma resposta não apropriada. Fica implícito, a partir das instruções nas provas, que o conhecimento prévio do candidato não deve interferir nas informações declaradas no texto. Conforme Hill e Parry (1992) argumentam, chega a ser irônico que os leitores “sejam encorajados a deixarem de lado toda a bagagem de conhecimento que é crucial para a efetiva interação deles com o texto”[23]. A título de ilustração, é apresentada a seguir uma tela com esse tipo de questão.

[23] No original: *readers are encouraged to hold separate the very knowledge which is crucial to their effective engagement with text* (p. 458).

Figura 6 – Tela TOEFL – exemplo de questão com base no texto

Na tentativa de controlar esses aspectos, os testes padronizados passam pelo crivo criterioso do procedimento de validação, que será apresentado a seguir, no que diz respeito a sua confiabilidade, validade, economia, facilidade de aplicação e relevância das normas e de sua aplicabilidade dos conteúdos para as características específicas do público a que se destina.

1.3.4 Validação de instrumentos de avaliação

> A validade é uma característica relativa e existe em diferentes graus. Sua problemática não está condicionada em termos de tudo ou nada. Um teste pode ser válido para certos fins e determinado grupo, mas não o ser para outros fins e indivíduos. A validade não é uma função do teste apenas, mas do uso que dele se faz. (VIANNA, 1976, p. 185)

Nesta seção é dada continuidade à revisão da literatura, no que tange aos testes de compreensão, enfatizando a validação dos instrumentos de avaliação. Primeiramente, são revisitados alguns dos mais importantes estudiosos da validação em testes – com foco especial nos testes de compreensão - trazendo à tona os mais destacados conceitos impressos ao termo *validação,* em uma perspectiva histórica.

1.3.4.1 Perspectiva histórica da validade de testes: a evolução do conceito

> Os testes de compreensão não são mais do que amostras de indicadores da leitura efetiva[24] (FARR e CAREY,1986)

A validação é o processo de examinar a precisão de uma determinada predição ou inferência realizada a partir dos resultados de um teste (RAYMUNDO, 2005). É, assim, o *olhar* atento sobre o *diagnóstico* para que ele possa fornecer um confiável *prognóstico*[25].

---

[24] No original: *All reading comprehension tests are essentially nothing more than samples of indicators of "real reading"*. (p. 33)

[25] Esses dois termos serão utilizados na seção de análise de dados e sustentam uma das hipóteses da presente pesquisa, daí a preocupação em deixar claro o sentido com que estão sendo empregados: De acordo com o dicionário Aurélio séc XXI, diagnóstico é *o conjunto dos dados em que se baseia um conhecimento ou determinação duma doença pelo(s) sintoma(s), sinal ou sinais e/ou mediante exames diversos* ao que se opõe a

Validação é a busca pela autenticação de um procedimento que faz parte das mais variadas esferas de inserção do indivíduo: a avaliação.

A ubiqüidade dos testes de compreensão sugere uma espécie de validade presumida e, de fato, muitos membros da comunidade acadêmica, bem como o público em geral, têm considerável fé em seus resultados. Mas se essa fé tem garantias, é uma questão a se considerar. Conforme Farr e Carey (1986), "a prática educacional convencional aceita os resultados dos testes de compreensão, independentemente do formato, como um índice válido acerca da habilidade de compreensão de uma criança, uma comunidade ou até mesmo de uma nação"[26].

O processo de validação não tem fim, ao contrário, pressupõe continuidade e deve ser repetido diversas vezes para o mesmo instrumento. Conforme Raymundo (2005), valida-se não propriamente o teste, mas a interpretação dos dados decorrentes de um procedimento específico.

Porém, nem sempre o fenômeno da validação foi visto sob esse ângulo, o que resultou, ao longo da história, em diferentes definições para o conceito. Conforme Tumolo (2005), a busca da validade já aparecia na preocupação que os filósofos tinham em validar tudo dentro do método científico. Messick (1989) aponta que filósofos como Leibnezi, Lock, Kant, Hegel e Singer já apresentavam em sua época seus diferentes modos de investigação considerando o processo de coleta de evidência de validade, ou seja, o processo de validação dentro do método científico.

Na arena específica dos testes, a validade aparece, segundo Tumolo (2005), por volta de 1920, época em que os elaboradores de testes começavam a buscar evidências de confiabilidade e validade em testes de inteligência nos EUA.

---

prognóstico, que é c*onjetura sobre o desenvolvimento de um negócio, de uma situação, etc.; predição; Juízo médico, baseado no diagnóstico e nas possibilidades terapêuticas, acerca da duração, evolução e termo de uma doença*. Trazendo para a arena da compreensão leitora, é preciso atenção especial ao se tomar o resultado de um teste específico (*diagnóstico*) como projeção generalizadora (*prognóstico*) da habilidade de leitura do candidato nas demais situações de uso real da língua.

[26] No original: *the conventional practice accepts the results of reading comprehension tests (regardless the format for the test) as a valid index of a child's, a community's, even a nation's ability to comprehend* (p. 33).

Conforme Shepard (1993), da década de 20 à década de 50 do século XX, a validade era definida como qualquer coisa com a qual um teste pudesse ser relacionado. Se o resultado de um teste fosse correlato com o resultado de um outro, então o primeiro teste era considerado válido.

Em 1954 surge o conceito da validade de construto como resultado de um empenho da Associação Psicológica Americana durante a elaboração do código de ética profissional (SHEPARD, 1993). Até o início da década de 60, segundo Messick (1989), a validade era considerada como sendo de quatro tipos: de conteúdo, preditiva, co-ocorrente e de construto. A validade de conteúdo, conforme definição do autor, refere-se às conclusões baseadas na avaliação de uma comparação entre o conteúdo do teste e as situações ou assuntos de interesse.

Na definição do *Dictionary of Language Testing* (DAVIES et al, 2002, p. 34) validade de conteúdo é uma validade conceitual ou não estatística, baseada em uma análise sistemática do conteúdo do teste para determinar se ele inclui uma amostra adequada do domínio alvo a ser medido. Uma amostra adequada envolve assegurar que todos os principais aspectos estão cobertos e em proporções apropriadas. Esse domínio alvo pode ser um módulo de ensino específico ou uma área específica de conhecimento, um domínio educacional ou ocupacional específico (para um teste de proficiência com propósito específico), ou a língua como um todo e seus possíveis usos. Ainda de acordo com o dicionário, na perspectiva teórica subjacente aos testes, o termo domínio "é utilizado de forma mais ampla para denotar a porção do universo total do tema (lingüístico ou outro) que está sendo testado ou sobre quais inferências serão retiradas"[27].

A validade preditiva refere-se, segundo Messick (1989), a quanto o teste pode prever sobre a performance ou desempenho do sujeito na situação de interesse. Ou ainda, mais tarde revisitando esse tipo de validade, Bachman (1990) diz que a validade preditiva requer a demonstração da relação entre os resultados de um teste e a performance real no critério. Conforme o *Dictionary of Language Testing* (DAVIES et al, 2002), "a validade preditiva mede o quanto um teste é capaz de prever sobre o desempenho do indivíduo em um critério

[27] No original: *The term is used more broadly to denote that portion of the total universe of subject matter (linguistic or other) which is being tested or about which inferences are to be drawn* (p. 49-50).

externo”[28]. Segundo o dicionário, uma vez que o principal propósito de um teste é fornecer informações sobre o comportamento provável do sujeito no mundo real, a predição do critério de performance, ou o aqui chamado *prognóstico*, é essencial para a validação do teste.

A validade preditiva assume particular importância em testes de proficiência em que o critério pode ser tão distante ou vago que o próprio teste pode ter que combinar os papéis de prognosticador e de critério. O critério, ainda segundo o dicionário, é uma variável externa que pode ser o conteúdo programático, o julgamento de um professor, a performance no mundo real ou outro teste com o qual o teste sob validação será comparado. Um teste busca fornecer um meio de representar o critério, uma vez que geralmente não é possível observar/ medir todos os elementos do critério. A performance no teste é usada para prever a performance do candidato no critério, que no caso dos testes de proficiência é a situação de uso da leitura no mundo real.

Segundo Messick (1989), a validade co-ocorrente refere-se a quanto um teste consegue estimar a atual condição do sujeito no critério de interesse. No *Dictionary of Language Testing* (DAVIES et al, 2002, p. 30), esse tipo de validade diz respeito à relação entre o que é medido por um teste (geralmente um teste recém desenvolvido) e outra medida de critério já existente, que pode ser um teste padronizado bem estabelecido, um conjunto de julgamentos ou alguma outra variável quantificável. Se as duas medidas funcionam semelhantemente, elas são consideradas como tendo validade convergente.

O problema da validade co-ocorrente, segundo o dicionário, é em relação à ênfase que se põe na medida de critério já existente. Se o critério é por si mesmo válido, então a necessidade do novo teste sob análise é questionável. Por outro lado, se o teste sob análise representa uma nova idéia, então haverá muito pouco proveito em validá-lo a partir de um teste já existente, que é presumivelmente considerado como uma medida imperfeita do relevante construto de habilidade. Se o novo teste quer medir a habilidade em inglês para um propósito específico, haverá pouco valor em comparar a performance nesse teste com as habilidades demonstradas em um teste de inglês geral (não-específico).

---

[28] No original: *Predictive validity measures how well a test predicts performance on an external criterion* (p. 149).

Messick (1989) define a validade de construto como a validade que se refere a uma investigação das qualidades medidas por um teste que são indicativas do construto. Em outras palavras, refere-se à demonstração de que o instrumento realmente mede o que se propõe medir. Conforme o *Dictionary of Language Testing* (DAVIES et al, 2002, p. 33), a validade de construto de um teste em língua é uma indicação de quanto o teste representa a teoria de aquisição de língua subjacente a ele. A validade de construto envolve uma investigação das qualidades que um teste mede, fornecendo, assim, uma base lógica para o teste.

O construto pode ser definido como uma habilidade ou um conjunto de habilidades que serão refletidas na performance do candidato através do teste, e sobre quais inferências podem ser feitas sobre essa performance com base no resultado do teste. Um construto é geralmente definido a partir de uma teoria, no caso dos testes em língua de uma teoria da língua. Assim, um teste representa uma operacionalização da teoria.

Ainda segundo o dicionário, existem dois aspectos na validade de construto: teórico e empírico. Ambos dizem respeito à produção de evidências ou argumentos que sustentam as inferências que são feitas sobre os candidatos com base em sua performance no teste. A validade de construto é tradicionalmente examinada ao determinar a relação entre o aspecto empírico (os padrões dos resultados de testes) e o teórico (conceitos explicativos propostos).

Conforme Raymundo (2005), a constatação da validade de construto resulta do acúmulo, por diferentes meios, de várias provas que precisam ser analisadas em todos os seus detalhes. Esse tipo de validação visa detectar, entre outros aspectos, quais as variáveis com as quais os resultados do teste se correlacionam, quais os tipos de itens que integram o teste, o grau de estabilidade dos resultados sob condições variadas e o grau de homogeneidade do teste (se mede um só ou vários traços), com vistas a ter elementos que possam esclarecer o significado do instrumento. Um procedimento cauteloso como esse se faz necessário em função do peso que o resultado, bem como as decisões com base nesse resultado, têm na vida do candidato.

Em 1961, Robert Lado dedica um capítulo de seu livro, intitulado *Language Testing*, à validade. Nesse capítulo, o autor define o conceito como sendo essencialmente uma questão de relevância. Se o teste mede o que se propõe medir é considerado válido. Segundo

Chapelle (1999), validade para o autor era uma característica do teste, um fenômeno tudo-ou-nada. Ou o teste era válido ou não era.

Na década de 70, a validade nos testes em língua começava a ser redefinida, uma vez que a tendência para a competência comunicativa levava os pesquisadores a "investigar questões sobre a validação de construto para testes de competência comunicativa"[29]. Em 1980, conforme Chapelle (1999), alguns pesquisadores passavam a investigar processos de validação sob uma outra perspectiva que não a do método de correlação. Nessa nova perspectiva, tentavam entender o método da testagem de hipótese, enquanto outros estudiosos sugeriam a inclusão de novas idéias como parte do conceito de validade, tais como as emoções, ética, o efeito no sistema de ensino e na sociedade – que será discutido posteriormente – e a idéia de que um teste é válido para um propósito específico.

Dando continuidade à retrospectiva histórica, em 1980, Messick inaugura um novo *olhar* para o conceito de validade, contribuindo, segundo Tumolo (2005), com a idéia de se considerarem valores e conseqüências sociais no conceito. Essas conseqüências sociais serão discutidas na última seção do presente capítulo. Messick sugere também que "o significado da medida e, conseqüentemente, sua validade de construto, deve ser sempre perseguido – não apenas para sustentar a interpretação do teste, mas também para justificar seu uso"[30].

Farr e Carey (1986), em seu livro *Reading: what can be measured?,* discutem a questão da validade, ressaltando que, de um modo geral, os testes se baseavam em amostras limitadas do comportamento do leitor sob condições também limitadas, com interferência de inúmeras variáveis e, na maioria das vezes, estariam avaliando competências lingüísticas isoladas.

Em 1989, Messick publica um artigo intitulado *Validity,* no qual define validade como "um julgamento avaliativo integrado sobre a extensão em que as evidências empíricas e bases lógicas teóricas sustentam a adequação e a apropriabilidade das inferências e ações baseadas nos resultados dos testes ou de outros modos de avaliação"[31].

---

[29] No original: "*probe questions about the construct validation for tests of communicative competence* (CHAPELLE, 1999, p. 255)

[30] No original: *the meaning of the measure, and hence its construct validity, must always be pursued – not only to support test interpretation, but also to justify its use* (MESSICK, 1989, p. 17).

[31] No original: *validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores*

Em resumo, a definição do conceito de validade apresenta três momentos distintos: no primeiro, a validade foi definida vagamente como qualquer coisa com a qual um teste pudesse ser correlacionado; no segundo momento, foram conceituados quatro tipos de validade; e o terceiro momento, que tem, conforme Tumolo (2005), grande influência nos testes atualmente, quando os pesquisadores passam a considerar a validade como única. No Quadro 5, a seguir, é possível comparar os conceitos antigos de validação – anteriores a 1980 - com os mais recentes, posteriores a esse período:

Passado	Presente
Validade era considerada como uma característica de um teste: o quanto um teste mede o que supõe medir.	Validade é considerada um argumento relativo à interpretação e uso do teste: o quanto as interpretações e usos do teste podem ser justificadas.
Confiabilidade era vista como uma condição distinta e necessária para a validade.	Confiabilidade pode ser vista como um tipo de evidência de validade.
Validade era frequentemente estabelecida através de correlações de um teste com outros testes.	Validade é debatida com base em uma série de bases lógicas e evidências, incluindo as conseqüências do teste.
Validade de construto era vista como um dos tipos de validade.	Validade é um conceito único com a validade de construto como central (evidência de conteúdo e relativa á critério podem ser usadas como validade de construto).
Estabelecer a validade era considerado como competência dos pesquisadores em testes responsáveis pelo desenvolvimento de testes de alto-risco e de grande escala.	Justificar a validade do uso do teste é responsabilidade de todos os usuários de testes.

Quadro 5 – Resumo dos contrastes entre conceitos de validação do passado e presente
Fonte: adaptação (CHAPELLE, 1999, p. 264)

Chapelle (1999) afirma que, embora seja possível ver as inúmeras modificações no conceito de validade a partir da contribuição de diversos estudiosos do tema, foi Bachman (1990), através de seu capítulo sobre validade, que mais influenciou os testes em língua, particularmente por definir a validação como um processo de demonstração de uma variedade de evidências sobre a interpretação e uso dos testes. Assim, entre os autores que contribuem para um olhar caleidoscópico ou prismático acerca desse assunto, destacamos em especial a perspectiva oferecida por Bachman (1990) e Bachman e Palmer (1996).

---

or other modes of assessment (MESSICK, 1989, p. 13).

1.3.4.2 O olhar de Bachman sobre a validade nos testes em língua

Em 1990, Bachman toma para si a tarefa de olhar as novas definições de validade e validação, surgidas na década de 80, à luz dos testes em língua, reaplicando-as nesse contexto específico. Conforme Tumolo (2005), o autor aceita algumas das inovações trazidas por Cronbach e Messick em 1980, como o fato de que o uso deve ser considerado para a validação e que o processo de validação requer todas as informações relevantes.

Segundo Bachman (1990), ao final do processo de validação, deveria ser possível responder à seguinte questão: até que ponto o teste analisado fornece uma base válida para fazer inferências sobre a habilidade de linguagem? Para o autor, o processo de validação deveria levar em consideração não apenas os indivíduos, mas também o contexto envolvido no uso da língua, incluindo a coleta de evidência como sustentação de um uso ou interpretação dada, a qual estará baseada em considerações lógicas, empíricas e éticas. A inovação que pode ser atribuída a Bachman é que ele divide as fontes de evidência para o processo de validação em base evidencial e base de conseqüência.

Em relação à base evidencial de validade, Bachman (1990) sugere que a evidência coletada para dar sustentação à decisão de um uso específico para um teste poderia ser agrupada em três tipos: a) relevância de conteúdo e cobertura de conteúdo; b) relação de critério e c) significação de construto.

A relevância de conteúdo refere-se a dois aspectos: 1) relevância do domínio (que no caso de testes em língua é a habilidade na língua) e 2) os métodos utilizados para elicitar a performance no candidato. A cobertura de conteúdo está relacionada com o quanto as tarefas solicitadas no teste representam adequadamente o domínio de comportamento em questão.

O segundo tipo de evidência requer identificar um critério apropriado e demonstrar que os resultados do teste estão funcionalmente relacionados a esse critério. O processo de validação pode demonstrar uma relação entre o resultado do teste e seu critério correspondente através dos procedimentos de validade co-ocorrente e de validade preditiva descritos na seção 1.4.4.1.

A significação de construto, o terceiro tipo de evidência indicada por Bachman, vem a ser a validade de construto renomeada pelo autor. É considerada por ele a mais fundamental entre os três tipos de evidência, uma vez que a validação de construto ocorre se os resultados do teste forem interpretados como indicadores de uma habilidade. Segundo Bachman (1990), a validação de construto pode ser comparada com um "caso especial de verificar, ou falsificar, uma teoria científica. Como uma teoria nunca pode ser 'provada', a validade de uso ou interpretação de qualquer teste está sempre sujeita à falsificação"[32].

A validação de construto, ainda segundo o autor, requer tanto uma análise lógica como investigação empírica. A análise lógica requer definir teórica e operacionalmente o construto, sugerindo que as hipóteses são feitas em ambos os níveis: teórico e operacional.

Bachman apresenta três tipos de evidência empírica para sustentar a validade de construto: correlacional (quantitativa), experimental (quantitativa) e análise de processo (qualitativa). A análise de processo permite a investigação do processo envolvido na performance das tarefas.

Outro aspecto considerado por Bachman (1990) como essencial para a base evidencial de validade de construto é a tendenciosidade do teste (*test bias*). A tendenciosidade, conforme o Dictionary of Language Testing (DAVIES et al, 2002), "é um fator ou grupo de fatores inerente a um teste que sistematicamente impede o acesso a uma estimativa válida acerca da habilidade dos candidatos"[33].

Ainda conforme o dicionário, um teste que possui validade de construto para candidatos de diferentes cenários lingüísticos, ou seja, um teste que mede a mesma habilidade ou grupo de habilidades da mesma forma para diferentes grupos pode fornecer um prognóstico tendencioso (*biased prediction*) acerca da performance em algum outro domínio.

Segundo exemplo do próprio dicionário, um teste de inglês pode ser igualmente válido como medida de proficiência (*diagnóstico*) tanto para imigrantes de longo prazo como para estudantes estrangeiros com visto de residência temporária. No entanto, como predição (*prognóstico*) da performance acadêmica futura desses sujeitos, como por exemplo, a

---

[32] No original: *a special case of verifying, or falsifying, a scientific theory, and just as a theory can never be 'proven', the validity of any given test use or interpretation is always subject to falsification* (p. 256).
[33] No original: *Is a factor or factors inherent within a test that systematically prevents access to valid estimates of candidates' ability* (p.15).

capacidade de aprender com a leitura, o teste pode ser menos eficaz para o último grupo devido ao conjunto de outros fatores sociais, que podem interferir nessa capacidade. Exemplos desses fatores seriam a habilidade de se aculturarem à nova realidade acadêmica e acesso a sistemas de apoio.

Em se tratando da base de conseqüência de validade, Bachman (1990) aponta que, ao considerar a validade dos resultados e usos dos testes, os elaboradores e usuários de testes devem considerar todas as questões éticas e políticas envolvidas. Segundo o autor, tais questões vão "além da demonstração científica de evidências lógicas e empíricas para a consideração das conseqüências potenciais dos testes....saindo dos confortáveis limites da lingüística aplicada e da teoria psicométrica para a arena da política pública"[34].

Assim, convergindo seu olhar com o de Messick (1989), Bachman (1990) destaca a necessidade de se olhar para as conseqüências que o uso dos resultados de um teste para um fim específico pode acarretar no sistema educacional e na sociedade, a fim de determinar se os benefícios são os esperados e se eles se sobrepõem às conseqüências negativas tanto para os indivíduos como para a sociedade.

A análise em torno da validação do teste deve levar em consideração um equilíbrio entre a teoria subjacente que determina os valores a serem implementados e que é operacionalizada pelo teste e o componente psicométrico que determina de que modo os valores serão implementados.

Para justificar o uso dos resultados de um teste como base para decisões sobre a vida futura do candidato, devem ser apresentadas ou discutidas evidências em relação à sua relevância e as conseqüências da decisão devem ser consideradas (TUMOLO, 2005). Essas evidências em relação à relevância dos resultados são sugeridas no modelo de utilidade dos testes proposto por Bachman e Palmer (1996) apresentado a seguir. O modelo será abordado com mais detalhes por dar conta, a nosso ver, de aspectos importantes a serem considerados na elaboração de testes para avaliação da compreensão que embasaram a presente investigação.

---

[34] No original: *beyond the scientific demonstration of empirical and logical evidence to the consideration of the potential consequences of testing....moving out of the comfortable confines of applied linguistics and psychometric theory and into the arena of the public policy* (BACHMAN, 1990, p. 281).

1.3.4.3 O olhar de Bachman e Palmer: o modelo de utilidade dos testes

Conforme os autores, ao utilizarem-se os resultados de um teste em língua, ou seja, o "diagnóstico", para fazer inferências sobre a habilidade que o indivíduo tem em uma determinada língua, ou seja, o "prognóstico" - e possivelmente, a partir disso, tomar vários tipos de decisões sobre sua vida futura – dever-se-ia poder demonstrar como a performance naquele teste em língua está relacionada com o uso da língua em situações específicas outras que não o teste em si. Assim, os autores baseiam seu modelo em dois princípios básicos: "a) a necessidade de uma correspondência entre a performance em testes em língua e o uso da língua; e b) uma definição clara e explícita das qualidades da utilidade do teste"[35].

A fim de se poder demonstrar essa relação, é necessária uma estrutura que possibilite tratar a performance em um teste em língua como uma instância específica do uso da língua. Para os autores, é necessária uma estrutura que permita utilizar as mesmas características para descrever o que se acredita serem características essenciais tanto para performance em teste em língua como para um outro uso da língua que não o de teste.

Assim, parte-se de um conceito geral, que é consenso entre muitos estudiosos de testes em língua (Vianna, 1989; Bachman, 1990; Bachman e Palmer, 1996; Messick, 1989; Chapelle, 1999), de que um teste tem uma utilidade específica, uma vez que é desenvolvido para avaliar uma determinada situação. Assim, um mesmo instrumento pode não ser capaz de medir, de modo globalizado, as diferentes dimensões do sujeito. Para ser útil, o teste deve ser elaborado com um propósito específico, para um grupo específico de candidatos e um domínio específico de uso da língua em mente.

Os autores referem-se a esse domínio como "domínio do uso da língua alvo ou domínio TLU (*Target Language Use*) e as tarefas nesse domínio como tarefas TLU[36]". O domínio do uso da língua alvo é um dos principais critérios que direcionam o olhar desta investigação para alguns aspectos de instrumentos padronizados de avaliação da compreensão como o TOEFL.

---

[35] No original: *The need for a correspondance between language test performance and language use; A clear and explicit definition of the qualities of test usefulness* (BACHMAN e PALMER, 1996, p. 9)
[36] No original: *We will refer to this domain as a target language use, or TLU domain, and the tasks in the TLU domain as TLU tasks* (BACHMAN e PALMER, 1996, p. 18).

Para ilustrar, consideremos que a situação específica para qual o teste se destina, que vem a ser o caso do TOEFL, seja a entrada de candidatos - cuja língua mãe não é o inglês - em uma universidade em que o inglês é a língua mediadora. Com base no modelo de utilidade, o teste deve ter autenticidade, ou seja, contemplar as situações reais de uso da língua que o candidato enfrentará em sua vida acadêmica, tais como as que envolvam as quatro habilidades fundamentais da língua – escrever, ler, falar e entender – em um padrão acadêmico/formal.

O teste deve dar conta de verificar se o candidato estará apto a se comunicar em inglês expressando seu pensamento e conseguindo entender o que lhe dizem; escrever empregando vocabulário e estrutura acadêmica; e, no caso específico da seção de leitura, ler de modo a reter informações e, conseqüentemente, aprender com a leitura, como terá que fazer em uma TLU.

A Figura 7, a seguir, representa esse critério da autenticidade: a relação entre performance no teste em língua e uso real da língua. Tal critério, segundo os autores, deve ser considerado para o *design*, desenvolvimento e uso dos testes em língua:

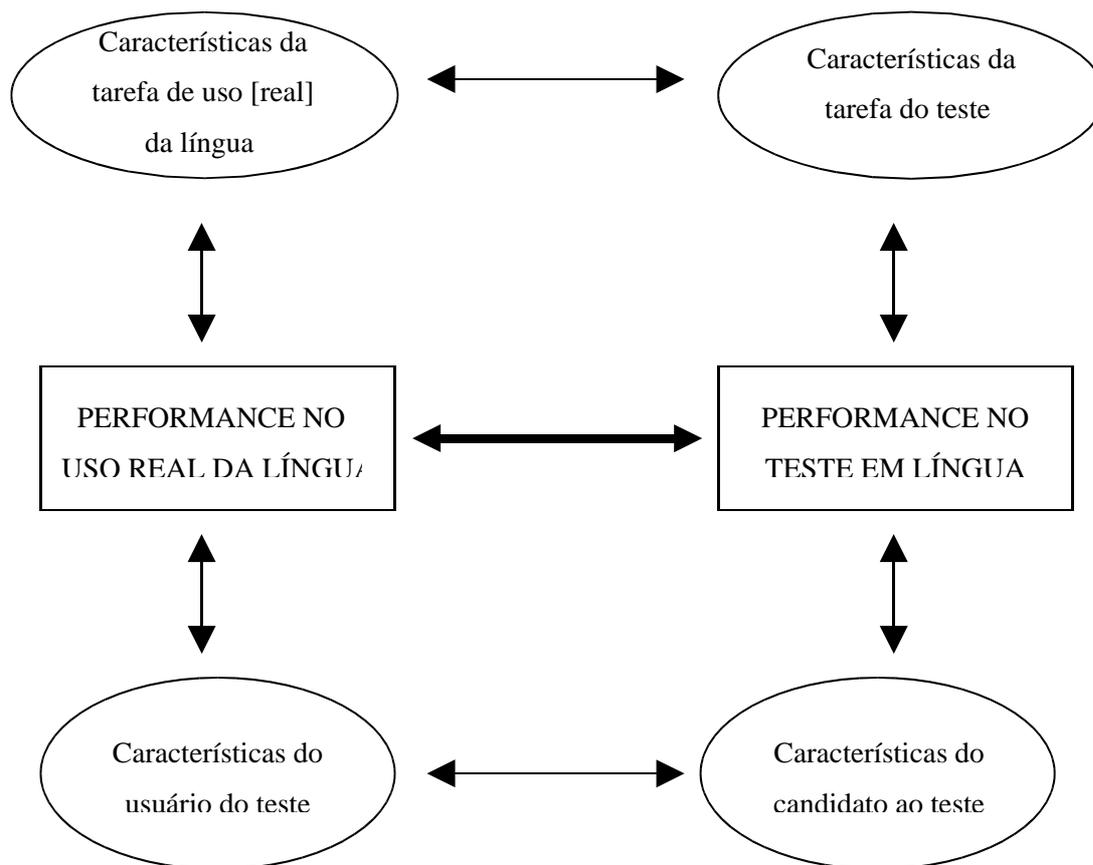



Figura 7 – Correspondência entre o uso da língua e a performance em teste em língua
Fonte: adaptação (BACHMAN e PALMER, 1996, p. 12)

Os autores sugerem que a autenticidade deve ser analisada sob diversas perspectivas e que todas elas devem ser levadas em consideração no desenvolvimento e uso dos testes em língua.

Contudo, é possível perceber que, por mais que as políticas de desenvolvimento de testes de compreensão persigam essa autenticidade, ela acaba não ocorrendo plenamente segundo Bachman e Palmer (1996) nem no sentido textual nem no sentido social. No sentido textual, para ilustrar, o texto que o candidato tem diante de si nem sempre é apresentado na íntegra. No TOEFL, como já referido na seção sobre a linguagem dos itens de múltipla-escolha, o texto sofre recortes e modificações. Se o texto for, na verdade, um excerto retirado de um texto maior e o leitor não tiver acesso a esse texto maior ou a dados como tipo do texto, título, autor, data de publicação, editora, suporte em que foi veiculado, entre outros, a passagem acaba tendo pouca semelhança com textos autênticos que o leitor vai encontrar em

situações ou tarefas de uso real da língua. Sendo assim, a performance desse leitor diante do excerto não revela como ele lidaria com esses dados presentes em textos nas situações reais, nem se ele saberia fazer uso dos mesmos para a construção de sentidos.

É exatamente a ausência de tais dados que pode comprometer a autenticidade no sentido social pois o sentido de um texto não depende apenas do que é declarado pelo autor, mas também das condições sociais e históricas sob as quais o texto foi produzido. O sentido social de um texto é constituído da intersecção entre os dados do texto, a bagagem que leitor traz e a localização específica no tempo e espaço no qual o texto foi produzido. Um mesmo texto pode ter sentidos sociais diferentes se lido em condições sociais, históricas e temporais diferentes, conforme apresentado nos aspectos cognitivos derivados do texto.

A interatividade, outro elemento presente no modelo de utilidade proposto por Bachman e Palmer (1996), é definida como o grau e o tipo de envolvimento das características individuais do candidato ao realizar uma tarefa do teste. Segundo os autores, as características individuais que são mais relevantes para os testes em língua são a habilidade do candidato na língua (conhecimento da língua e competência estratégica ou estratégias metacognitivas), conhecimento local e o esquema afetivo do indivíduo.

Fica evidente no olhar dos autores que tanto a autenticidade quanto a interatividade são qualidades relativas, não podendo assim se atribuir a um teste a sentença de que *é* ou *não é* autêntico ou interativo. Reforçam a perspectiva de que a validação dos instrumentos de avaliação não é um fenômeno tudo-ou-nada. Os autores ainda salientam que a autenticidade, interatividade e validade de construto dependem todas de como se define o construto ‘habilidade na língua” para uma determinada situação de teste.

Outra qualidade, conforme o modelo de utilidade, é a praticidade. Enquanto as demais qualidades pertenciam aos usos que são feitos dos resultados dos testes, a praticidade pertence principalmente às formas nas quais o teste será implementado e se ele será desenvolvido e utilizado de fato. Por exemplo, se os recursos necessários para a implementação excederem os recursos financeiros disponíveis, o teste será considerado não-prático. Porém, isso não significa que essa qualidade seja menos importante que as demais.

Um teste considerado prático é um teste cujo *design*, desenvolvimento e uso não requerem mais recursos do que os que estão disponíveis. Segundo os autores, vários tipos de recursos podem ser identificados: recurso humano, recurso material e tempo. Os recursos específicos solicitados irão variar de uma situação para outra, assim como os recursos disponíveis. Assim, a praticidade, como a validade de um modo geral, apenas pode ser determinada para uma situação específica e a partir de um determinado ângulo de visão.

A última qualidade a ser considerada, pertencente ao modelo de Bachman e Palmer (1996), é o impacto dos testes na sociedade e nos sistemas educacionais, bem como nos indivíduos que fazem parte desses sistemas. Nessa perspectiva, o uso dos testes impacta em dois níveis: um nível micro, referente aos indivíduos que são afetados pelo uso do teste, e um nível macro, referente ao sistema educacional ou à sociedade. A característica do impacto pode ser representada através da Figura 8:

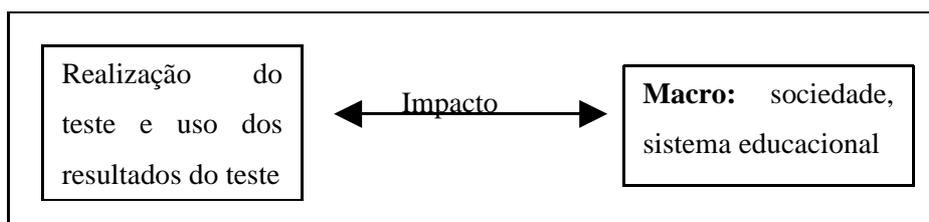



Figura 8 – Impacto dos testes
Fonte: adaptação (BACHMAN e PALMER, 1996, p. 30)

Segundo os autores, os atos de administrar e aplicar um teste que revelam certos valores e objetivos de uma teoria subjacente têm suas conseqüências. Da mesma forma, os usos que são feitos dos resultados dos testes sugerem valores e objetivos próprios e também acabam tendo conseqüências. Entre os aspectos que impactam na sociedade, sistema educacional e indivíduo e que têm sido observados com grande interesse por vários estudiosos e já brevemente mencionado no presente capítulo, está o chamado efeito *washback*[37].

O efeito *washback*, na perspectiva dos autores, vem sendo amplamente apontado na arena dos testes em língua como o impacto direto dos testes sobre os indivíduos. Hughes (1989), citado por Bachman e Palmer (1996), define-o como o efeito dos testes no processo de ensino e aprendizagem e argumenta que os testes podem tanto ter um efeito benéfico

[37] Optamos por manter o termo em inglês por não encontrarmos uma palavra em português que desse conta de seu sentido e por esse ser um termo técnico encontrado na literatura internacional.

como prejudicial sobre esse processo. Cohen (1994) discute os efeitos desse fenômeno de forma mais generalizadora, pensando em como os instrumentos de avaliação afetam as práticas e crenças educacionais.

Wall e Alderson (1993), a partir de um ângulo um pouco diferente, defendem veementemente, com base na extensa pesquisa empírica que vem sendo feito na área, que os elaboradores e usuários de testes não podem simplesmente admitir que os testes causem um impacto no ensino e na aprendizagem, mas que deveriam, na verdade, investigar as áreas específicas, a direção (positiva ou negativa) e a extensão do suposto impacto.

De acordo com o *Dictionary of Language Testing* (2002), o efeito *washback* é o "efeito da testagem no ensino. Esse efeito pode ser tanto positivo quanto negativo. O efeito negativo ocorre quando os itens do teste são baseados em uma visão antiquada da língua, o que mantém pouca relação com o currículo do ensino"[38]. Conforme exemplo no dicionário, se a habilidade da escrita é testada em um determinado instrumento de avaliação, somente através de questões de múltipla escolha, então há uma grande pressão sobre o ensino para a prática de tais itens ao invés de praticar a habilidade da escrita em si. Qualquer teste importante, no caso da compreensão leitora, que permaneça muito baseado em estrutura pode ser apontado como tendo efeito negativo se sua influência evita as modificações no ensino da proficiência em direção a uma metodologia mais comunicativa.

Segundo Peirce (1992), o efeito *washback* de um teste, algumas vezes referido como validade sistêmica de um teste, se refere ao impacto de um teste sobre a pedagogia escolar, desenvolvimento do currículo e política educacional. A autora aponta os resultados de pesquisas (Shohamy, 1992) demonstrando que a inserção, em âmbito nacional, de três testes em língua em Israel teve um impacto dramático na pedagogia escolar e na política educacional do país[39].

Em relação ao TOEFL, a autora aponta como um dos indicadores de seu efeito a vibrante indústria de material preparatório para o teste. Ela sugere que na fase de revisão e validação do teste seja considerado, pela equipe de elaboradores, que tipo de impacto o

[38] No original: *it is the effect of testing on instruction. Language test washback is said to be either positive or negative. Negative washback is said to occur when test items are based on an outdated view of language which bears little relationship to the teaching curriculum* (p. 225).

[39] Um outro exemplo desse efeito, muito comum no Brasil, é o do ensino que visa de maneira específica a preparação para o vestibular.

TOEFL tem tido na pedagogia de sala-de-aula e na política educacional, não apenas na América do Norte mas também em alguns dos diversos países nos quais o TOEFL é administrado.

Com interesse comercial de fazer com que seus "clientes" obtenham sucesso, grande parte dos materiais didáticos elaborados por essa indústria – e que acabam servindo de parâmetro nos cursos preparatórios – visa fazer o aluno atingir uma pontuação elevada através de manuais e testes simulados. Para ilustrar, é apresentada a seguir a chamada em um material preparatório que foi o utilizado em nossa pesquisa. O material é composto de um livro e um *cd* com uma simulação do teste e traz em sua capa a seguinte chamada:

> A segunda edição do TOEFL CBT com CD-ROM da Kaplan combina uma completa experiência prática de teste baseado no computador com estratégias para ajudar você a aumentar sua pontuação. Essa poderosa combinação torna esse livro uma ferramenta altamente eficiente para ajudar você a atingir uma pontuação mais alta no TOEFL CBT e tornar você mais competitivo para a seleção em uma universidade americana. Pontuação mais alta garantida[40]! (Kaplan TOEFL CBT Exam. Test Preparation).

Em momento algum dessa chamada é prometido tornar o comprador do material um melhor leitor, mas sim um melhor realizador do teste. Essa influência dos testes ao servirem de parâmetro para o ensino, como ocorre em muitos casos com o TOEFL, pode ser perigosa quando resulta em um sistema de ensino que não contemple a leitura com o propósito de aprendizagem, mas apenas como um jogo de "adivinhação", apenas para acertar o teste, podendo acabar por gerar um vazio após a leitura.

Na visão de Peirce (1992), os professores de inglês enquanto língua estrangeira (*English as a Second Language*), doravante ESL, deveriam ser consultados para determinarem que construto de leitura deveria ser medido no teste de leitura do TOEFL e como o teste pode melhor servir os interesses de seus programas e as necessidades de seus alunos. Os candidatos também deveriam ser consultados para determinar como a preparação para o TOEFL poderia promover o aprendizado de fato da língua e como a ansiedade associada com a situação de teste do TOEFL (evento social) poderia ser aliviada.

---

[40] No original: *Kaplan's TOEFL CBT Exam with CD-ROM, Second Edition combines a complete computer-based practice test experience plus strategies to help you increase your score. This powerful combination makes this book a highly effective tool to help you score higher on the TOEFL CBT and to make yourself more competitive for admission in an American university. Higher score Guaranteed!*

Assim, esse referencial teórico forneceu bases para a compensação da perturbação em relação ao fenômeno da compreensão leitora e de sua avaliação. Segundo Kastrup (1999), a compensação se dá quando o organismo consegue atribuir significado a uma perturbação. Perturbar, conforme a autora, significa afetar, colocar problema. Tudo começa com perturbações. O meio comparece perturbando e não transmitindo informações e é essa perturbação gerada pela interação com o meio que gera a cognição. Com base nessa revisão da literatura e de posse da perturbação inicial foi executada a pesquisa que será descrita no capítulo seguinte.

2 EM BUSCA DE UMA LENTE METODOLÓGICA ADEQUADA À PESQUISA SOBRE AVALIAÇÃO DA COMPREENSÃO LEITORA

> Fico intrigado à medida que redescubro continuamente que aquilo que parece ocorrer tão sem esforços na natureza pode ser tão trabalhoso para computar.[41] (HAYES, 1996)

A percepção de que nem sempre a performance em um teste em língua está relacionada com o uso da língua em situações específicas outras que não o teste *per se*, que vem a ser o critério de autenticidade já referido, nos fez olhar para a avaliação da compreensão leitora com uma inquietação pulsante. Somado a isso está o fato de que alguns instrumentos para avaliação da compreensão em leitura enfatizam capacidades cognitivas isoladas, como a memória de trabalho dos candidatos, enquanto se propõem a avaliar a compreensão como um todo.

Essas inquietações serviram como motivação para pensar acerca da validade preditiva da avaliação da compreensão. Até que ponto o *diagnóstico* de um leitor em um teste nos permite fazer um *prognóstico* de sua competência leitora? Ou ainda, quanto o resultado (percentual de acertos) de um teste se correlaciona com o nível de representação do texto? Cumpre lembrar que, na perspectiva aqui adotada, a avaliação da compreensão revela a concepção de compreensão subjacente a ela.

Uma avaliação de compreensão leitora, que tem por propósito a seleção de candidatos para o ingresso em universidades de língua inglesa, por exemplo, a fim de garantir essa autenticidade, deve contemplar tarefas que se assemelhem às tarefas que o candidato enfrentará na universidade posteriormente. Esse vem a ser um ideal perseguido por elaboradores de testes de compreensão leitora, dentre eles o TOEFL, que tem sua seção de leitura como objeto de estudo da presente investigação. Segundo o boletim do instituto que prepara as provas, o TOEFL tem essa autenticidade garantida, uma vez que, como já citado,

[41] No original: *I am intrigued as I continually rediscover that what looks effortless in nature can be so laborious to compute*. (Brian Hayes em um artigo publicado na American Scientist em Julho de 1996, citado em Kintsch (1998, p.247)).

“mede a habilidade de falantes não nativos de inglês em usarem e entenderem a língua inglesa da forma em que é falada, escrita e ouvida em ambientes de faculdades e universidades Norte Americanas”[42].

Assim sendo, uma das tarefas eleitas aqui como essenciais em ambientes de faculdades e universidades, devendo por isso estar contemplada nos testes, é a utilização da leitura como dispositivo para a aprendizagem. Essa tarefa é destacada tendo em vista o fato de que no contexto acadêmico - que é o contexto de destino desses calouros – a leitura será uma das principais fontes de aprendizagem. Para que haja essa aprendizagem é necessário que o leitor consiga fazer uma representação mental do texto, a qual é vista por Salles (2002) como um aspecto cognitivo necessário para a compreensão leitora, uma vez que é por meio dela que o leitor consegue a retenção da informação e a conseqüente aprendizagem.

Tal representação, que se dá aparentemente de uma forma tão automática e sem esforços em leitores “competentes”, tem inquietado muitos pesquisadores. Como se constrói o modelo de situação proposto por Kintsch e van Dijk (1978) na mente do leitor? Como verificar se isso acontece ou não? Nesse espectro, tem se apresentado como um desafio o estabelecimento de níveis os quais permitam vislumbrar com maior nitidez as representações de sentido que os leitores constroem durante a compreensão. Tal conhecimento é útil não apenas para se entender melhor o fenômeno da cognição, mas também para intervir pedagogicamente nos casos de leitores “não-competentes”. Graesser e colaboradores (1997, p.169) afirmam que um desafio persistente tem sido planejar tarefas experimentais que isolem os níveis de representação separados.

Assim, na presente pesquisa tomou-se como *lente de longo alcance* o modelo cognitivo de compreensão de textos de Kintsch e van Dijk (1978) e como *lente de curto alcance* o modelo de utilidade de Bachman e Palmer (1996) para se olhar para a validade preditiva da avaliação da compreensão leitora.

---

[42] No original: ***Reading** measures the ability to understand short passages similar in topic and style to academic texts used in North American colleges and universities. (*Information and *R*egistration Bulletin for TOEFL computer-based testing (TOEFL cbt), TOEFL paper-based testing (TOEFL pbt), test of spoken English (tse), 2006-07, p. 4)

O objetivo da presente pesquisa, através desses óculos com lentes progressivas, foi investigar o nível ou níveis de representação alcançados pelos sujeitos ao lerem um texto do TOEFL, e se esses níveis eram suficientes para executarem as tarefas do mesmo. Além disso, como um desdobramento desse objetivo, se buscou fazer uma comparação entre a abordagem ao texto feita pelo instrumento e o nível de compreensão resultante alcançado pelo sujeito. Para isso foi utilizada como contraponto uma abordagem com base na taxonomia de Barret (1968), apresentada no capítulo anterior, que serviu como uma proposta alternativa para a avaliação tradicional.

As seguintes questões nortearam o delineamento e encadeamento da pesquisa:

1. O que é compreensão textual?
2. Como fica a qualidade de leitura na avaliação da compreensão?
3. Até que ponto o critério de profundidade é levado em consideração nos testes?
4. Como é observada a relação entre diagnóstico e prognóstico?
5. Como é observado o modelo de utilidade?
6. Como é observado o critério de autenticidade?
7. Que conseqüências o modelo de compreensão em leitura adotado por um teste padronizado como o TOEFL traz para a relação ensino-aprendizagem e para a sociedade?

2.1 Os objetivos da pesquisa

O objetivo geral da presente pesquisa foi o de investigar se o nível ou níveis de representação textual alcançados pelos sujeitos, ao lerem um texto de um instrumento padronizado de avaliação, são suficientes para executarem as tarefas do teste. Para isso foram considerados os aspectos cognitivos fundamentais envolvidos na compreensão em leitura e no processamento da linguagem tendo em vista que compreender um texto significa formar uma estrutura mental que represente o sentido do texto (KINTSCH, 1998).

Entre os objetivos específicos da pesquisa estão:

- Verificar qual o nível de representação textual alcançado pelo sujeito após a realização de um teste padronizado de avaliação da compreensão em leitura e a correlação do mesmo com o escore obtido pelo candidato;
- Comparar a abordagem ao texto feita pelo instrumento e o nível de representação alcançado pelo sujeito;
- Identificar os principais constituintes da habilidade em leitura a serem contemplados nos testes de avaliação para garantir a autenticidade dos mesmos;
- Investigar as conseqüências do efeito *washback* conforme seção 1.4.4.3;
- Fornecer subsídios para a elaboração de propostas de instrumentos de avaliação da compreensão em leitura que evitem a formação indireta de um comportamento padrão de leitura.

2.2 Hipóteses

A primeira hipótese é de que o nível de compreensão textual suficiente para a realização das tarefas do teste TOEFL não é o nível ideal para uma representação mental do texto viabilizadora da aprendizagem.

Uma segunda hipótese, que vem a ser um desdobramento da primeira, é que, dependendo do grau de profundidade da abordagem ao texto, o nível de representação será maior ou menor.

2.3 Justificativa

Conforme a lente de Koda (2005), a tendência que prevalece na avaliação da compreensão em L2 é focalizar mais a proficiência da linguagem do que a compreensão leitora propriamente dita e pouco se tem feito para construir medidas fidedignas que atestem a habilidade em leitura de L2. Algumas justificativas para essa exigüidade: a) carência de modelos viáveis de leitura em L2 e de fundamentações teóricas testadas, uma vez que os maiores constituintes da habilidade em leitura em L2 permanecem incertos; b) os primeiros estudos sobre compreensão em leitura de L2 eram fortemente influenciados pela visão “universal” de leitura que afirmava serem as mesmas, mesmo entre línguas diferentes, as

habilidades necessárias para a compreensão, o que levou à crença de que todas as dificuldades experienciadas pelos leitores de L2 são provenientes de conhecimento lingüístico inadequado.

Somado a isso, outro fator motivacional é que alguns instrumentos de avaliação verificam capacidades isoladas dos sujeitos, tais como a memória de trabalho, além de, muitas vezes, não contemplarem, talvez em virtude da forma como são aplicados, importantes aspectos que fazem parte das dimensões cognitivas da compreensão em leitura.

A seção de leitura do TOEFL, na forma como é aplicada, com uma exigüidade de tempo, pode estimular a adivinhação das respostas e fazer com que o aluno, objetivando o escore, tenha um comportamento de leitura caracterizado por uma relativa superficialidade na interpretação. O que é verificado ao término da prova é uma habilidade de raciocínio rápido e capacidade de adivinhação, que são características necessárias, porém nem sempre suficientes para a compreensão textual e conseqüentemente para a aprendizagem com a leitura.

Com o desenvolvimento de um instrumento de avaliação alternativo, baseado na Taxonomia de Barret, procurou-se olhar para a compreensão com uma outra lente. Através dessa, buscou-se verificar a compreensão do texto em maior profundidade e contemplar as dimensões cognitivas da compreensão leitora que podem oportunizar ao candidato formar uma estrutura mental que represente o sentido do texto e desenvolver assim, um novo comportamento de leitura.

2.4 Método

Este é um estudo transversal, do tipo descritivo, com delineamento correlacional e comparativo (entre grupos) (ROBSON, 1993).

O projeto de pesquisa foi aprovado pelo Comitê de Ética em Pesquisa da Universidade de Santa Cruz do Sul, cadastro número 1483/06 (ANEXO A). Seguindo as orientações desse comitê, todos os participantes assinaram o Termo de Consentimento Livre e Esclarecido (ANEXO B).

Serão descritas a seguir, as etapas da produção[43] de dados, entre elas estão: a seleção de sujeitos, os instrumentos utilizados e os procedimentos adotados.

2.4.1 Seleção dos sujeitos da pesquisa

A amostra da presente pesquisa foi composta por alunos do curso de Letras e de cursos avançados de conversação do Centro de Línguas e Culturas (CELINC) da Universidade de Santa Cruz do Sul/RS.

Como no presente estudo são enfocados os níveis de representação textual relacionados aos escores das avaliações da compreensão, foram selecionados determinados níveis dos cursos citados a fim de se obter um grupo de participantes com proficiência leitora homogênea e de nível elevado. Tais níveis correspondem ao VIII semestre do curso de Letras e dos seguintes cursos do CELINC: Língua Inglesa para professores do departamento de Química e Física nível VII; Conversação em Língua Inglesa; Let's speak English VII e Inglês Instrumental para a área da Saúde VI. Para ingressar nesses cursos os alunos passaram por provas de nivelamento utilizadas pelo CELINC. O nível de proficiência nesses cursos é equivalente ao VII semestre do curso de Letras, uma vez que até o livro utilizado em alguns é o mesmo. O objetivo da observação de tais critérios de seleção foi o de que os participantes estivessem emparelhados, tanto quanto possível, em relação aos conhecimentos que possuem da língua inglesa bem como quanto à sua proficiência leitora na mesma. Os participantes são todos adultos, estudantes da graduação ou professores da universidade.

A amostra foi inicialmente composta de vinte e seis sujeitos que participaram de todas as etapas da testagem. Porém, em função da escolha de ter todo o experimento mediado pelo computador e *internet*, houve alguns problemas de ordem técnica, como a perda, devida ao não registro no sistema do ambiente de educação a distância da universidade, de algumas respostas dadas. Essas respostas faziam parte da análise qualitativa da investigação, e como eram necessários os dados de cada sujeito em todas as etapas da testagem, abandonamos os

[43] Optou-se aqui por utilizar o termo *produção* ao invés de *coleta* de dados por vermos no termo coleta a conotação de algo já pronto que será apenas colhido, coletado. Dessa forma, o conceito estaria indo de encontro à perspectiva da presente pesquisa que vê o leitor como *construtor/produtor* de uma estrutura de representação textual que demanda muito empenho e esforço por parte dele. Não *coletamos* algo pronto, mas sim trouxemos para a pesquisa dados que foram *produzidos* por nossos participantes.

dados desses, substituindo-os por outros sujeitos. Portanto, a amostra para fins de análise é composta pelos dados de vinte participantes.

2.4.2 Instrumentos de *produção* de dados

Para a produção dos dados desta pesquisa foram utilizados três instrumentos. Os dois primeiros constituem-se em testes de compreensão leitora tendo como suporte o computador. O terceiro instrumento utiliza a estratégia do reconto. Os instrumentos serão apresentados a seguir.

2.4.2.1 TOEFL

Nesta pesquisa foi utilizada uma, entre as várias opções de material preparatório para o TOEFL com testes simulados que se encontram disponíveis no mercado. Como a presente pesquisa enfoca a avaliação da compreensão leitora, apenas a seção de leitura do TOEFL foi empregada. A versão do teste aqui utilizada foi citada na seção 1.4.3.2 deste estudo.

2.4.2.2 OUR TOOL

Com a necessidade de um outro olhar sobre a avaliação da compreensão para que fosse feito um contraponto com a concepção TOEFL de leitura e, assim, verificar características de uma e de outra, lançou-se mão de uma perspectiva diferenciada de avaliação. A avaliação taxonômica, referida na seção 1.4.3.2, apresenta-se assim como uma proposta alternativa para a avaliação tradicional da compreensão em leitura empregada como uma lista de habilidades.

Assim, com base na Taxonomia de Barret (1968), a qual contempla as dimensões cognitivas e afetivas da compreensão em leitura, foi elaborado um instrumento de avaliação da compreensão, o qual foi carinhosamente batizado de OUR TOOL (Nossa Ferramenta), uma vez que sua construção se deu por várias mãos e sob alguns olhares atentos. Para efeito de padronização, já que a sigla do outro teste está em inglês, será utilizado doravante OUR TOOL para fazer referência ao instrumento desenvolvido na presente pesquisa.

Na tentativa de controlar variáveis intervenientes, optou-se por oferecer o novo instrumento de avaliação no mesmo suporte em que o TOEFL é apresentado, ou seja, no computador. Para isso, foi desenvolvido um sistema *web*, escrito em linguagem de programação específica para *intenet* (PHP), no qual o instrumento pode ser acessado, realizado e automaticamente corrigido. O instrumento está disponível no *site* http://mvb.viavale.com.br/karen, e também terá algumas de suas telas apresentadas ao longo do texto para auxiliar na visualização.

A forma em que o instrumento foi elaborado forneceu dados para uma posterior análise do processo de leitura do candidato, além da possibilidade de uma comparação equilibrada com o TOEFL. Entre os dados gerados pela OUR TOOL estão o tempo total que o candidato levou para realizar a prova, o tempo que levou em cada questão e que opção de resposta fez, como pode ser visto na tela a seguir:

Figura 9– Tela de correção da OUR TOOL

Nas instruções de resposta, em algumas questões, havia mais de uma alternativa correta, como pode ser visto na Figura 10 a seguir. Cada questão foi cuidadosamente redigida em função da correção automática realizada pelo computador que foi implementada na OUR TOOL e da não mediação de um instrutor/professor ao longo da realização da tarefa.

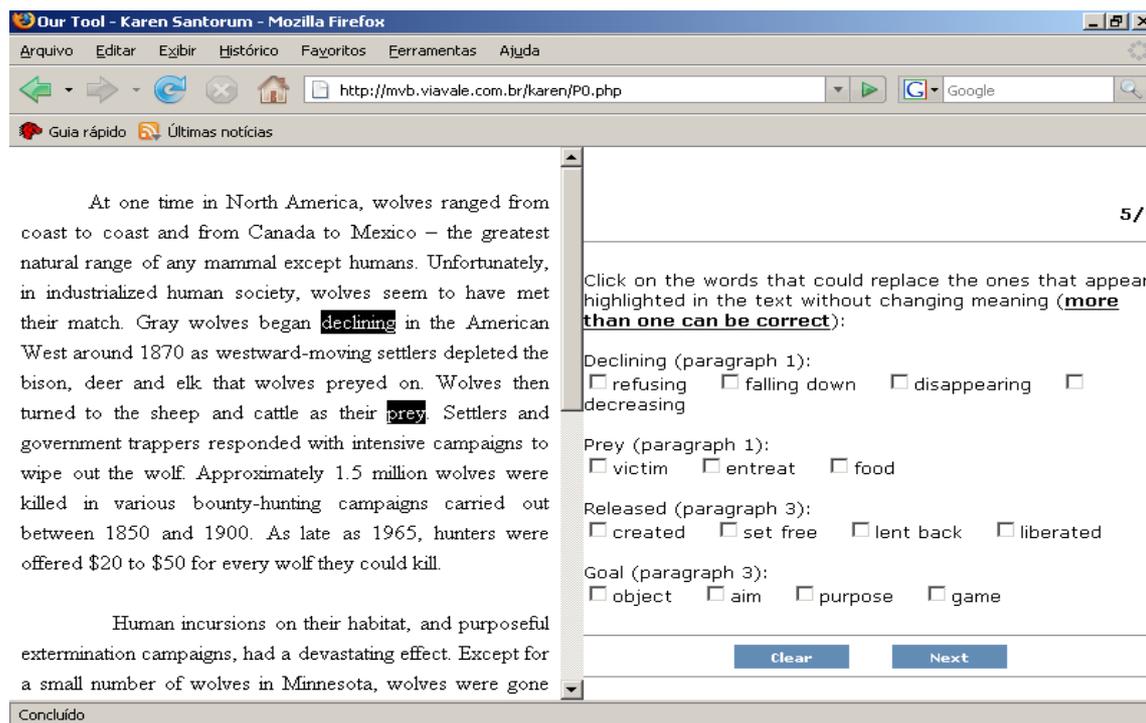



Figura 10 – Tela OUR TOOL exemplo de questão com mais de uma resposta correta

Em situações em que mais de uma alternativa estava correta, a instrução dada à ferramenta foi que seria considerado correto apenas se o candidato marcasse todas as respostas corretas. O candidato que marcasse apenas uma opção teria a questão considerada errada. A ferramenta foi construída de forma que novas questões pudessem ser cadastradas, para uso futuro da mesma como prática de leitura em sala de aula.

Assim como no TOEFL, em OUR TOOL os candidatos têm na tela um texto que fica aparente durante toda a realização da prova. Ao lado do texto vão aparecendo as questões que o candidato responde clicando com o *mouse* nas alternativas corretas, ou, outro diferencial em OUR TOOL, justificando sua resposta, digitando-a em um local específico, como mostrado na Figura 11 a seguir.

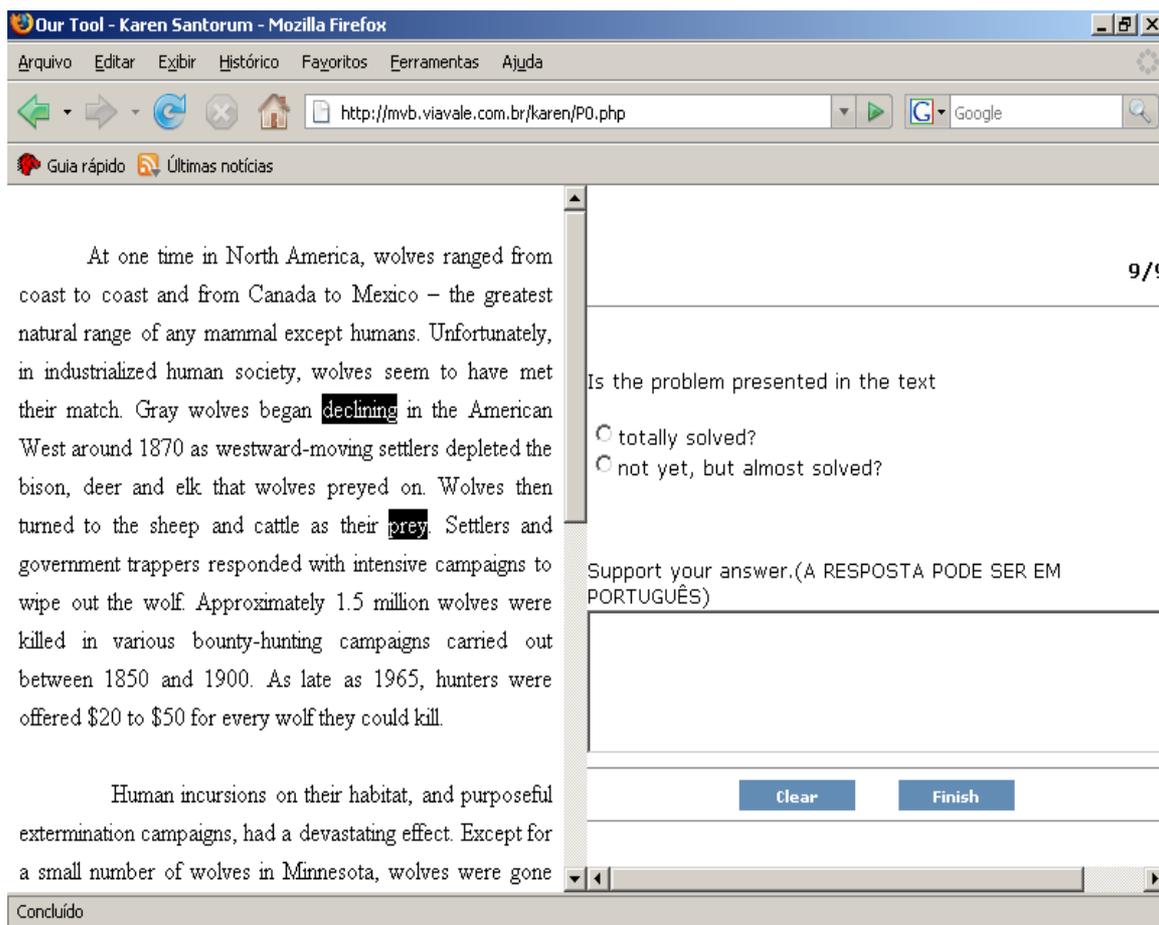



Figura 11 – Tela questão justificada OUR TOOL

Outra providência tomada no intuito de controlar as variáveis foi trabalhar o mesmo texto da versão do TOEFL utilizada para não correr o risco de ter textos com graus diferentes de dificuldade, uma vez que o foco estava centrado nas abordagens ao texto feitas pelos dois instrumentos. O texto utilizado em ambos instrumentos pode ser visto na íntegra no ANEXO C.

A taxonomia de Barret (1968), originalmente aplicada a textos narrativos, foi aqui adaptada a um texto não-narrativo. A seguir são dados exemplos de questões nas quais foram aplicadas explicitamente as categorias taxonômicas de Barret ao texto. Com o propósito de ilustração, estão reproduzidas aqui as telas como no instrumento original.

Questão 1 – Nível taxonômico: compreensão literal, reconhecimento de detalhes. Tipo de pergunta: múltipla escolha. Intencionalmente foi utilizado na questão um vocabulário diferente do que o apresentado no texto para evitar que o candidato respondesse baseado

apenas na identificação da palavra. *Quando a população de lobos cinza começou a decair no ocidente americano?*

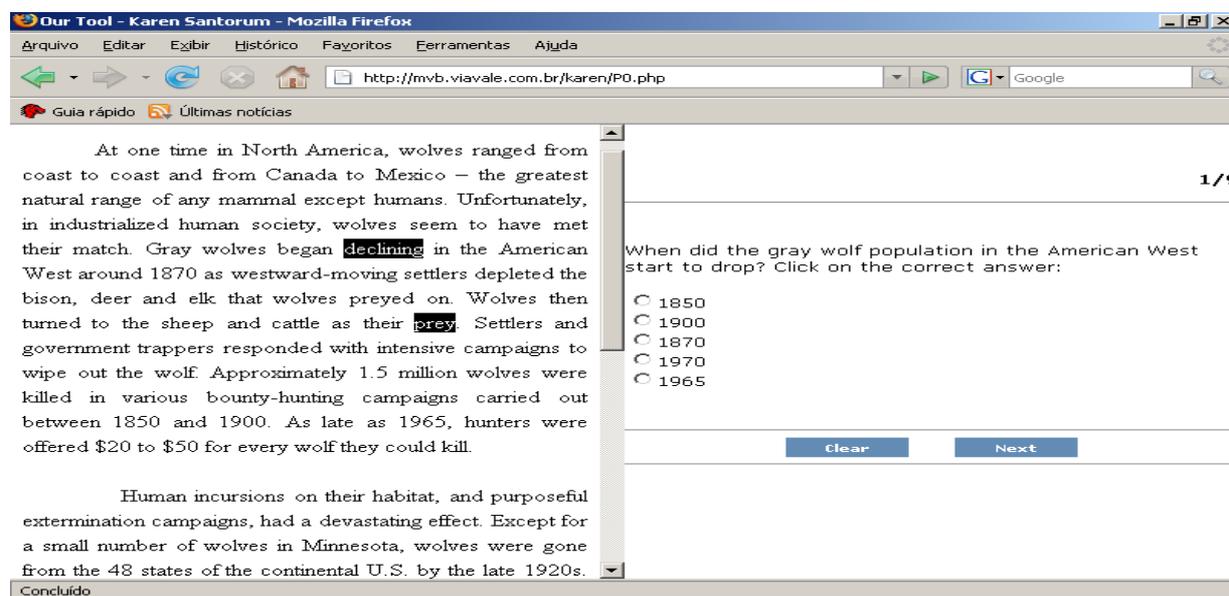



Figura 12 – Tela OUR TOOL questão 1

Questão 2: Nível taxonômico: compreensão literal, reconhecimento de seqüências. Tipo de pergunta: múltipla escolha. O candidato precisa identificar a ordem de incidentes ou ações apresentadas no texto através de um vocabulário semelhante porém não idêntico ao texto. *Clique na resposta correta.*

Figura 13 – Tela OUR TOOL questão 2.

Questão 5: Nível taxonômico: compreensão inferencial, inferência de detalhes. Tipo de pergunta: múltipla escolha com várias alternativas corretas. *Clique nas palavras que poderiam substituir as que aparecem destacadas no texto, sem mudar o significado:*

Figura 14 – Tela OUR TOOL questão 5

Questão 4: Nível taxonômico: compreensão inferencial, inferência de idéias principais. Tipo de pergunta: múltipla escolha. *Qual o objetivo principal do texto?*

Figura 15 – Tela OUR TOOL questão 4

Questão 6: Nível taxonômico: compreensão inferencial, inferência de causa e efeito. Tipo de pergunta: múltipla escolha. Resposta não explícita no texto. *Por que foi oferecido dinheiro aos caçadores por cada lobo que eles conseguissem matar?Clique na melhor resposta.*

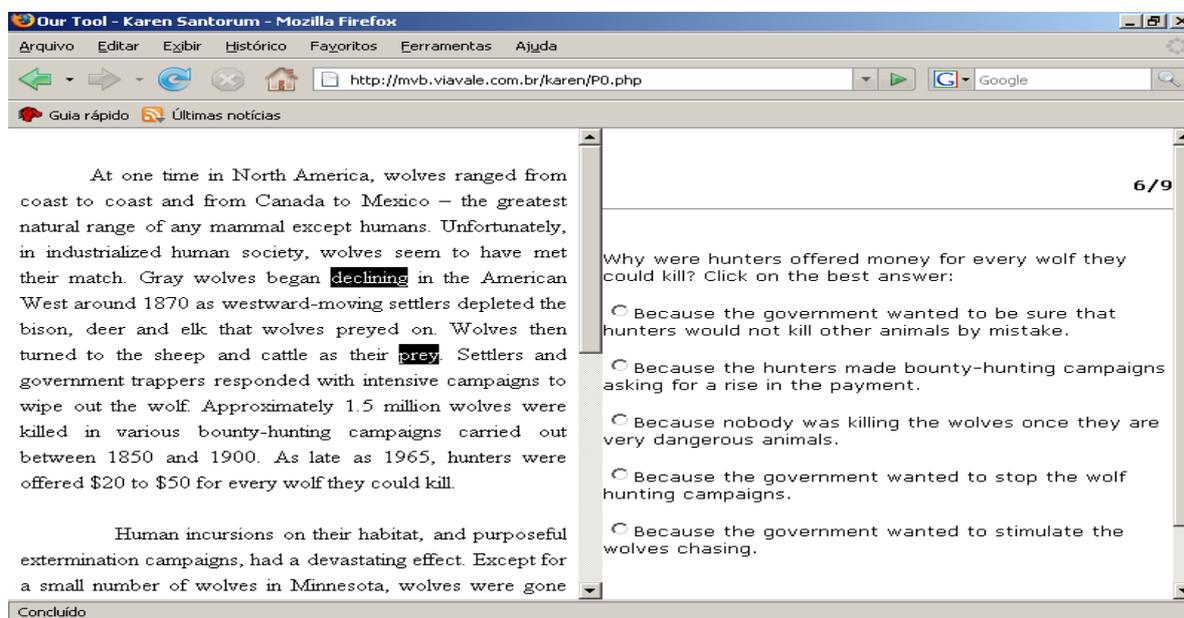



Figura 16 – Tela OUR TOOL questão 6

Questão 7: Nível taxonômico: leitura crítica, julgamento de valores. Nível taxonômico: avaliação, tenta avaliar o impacto psicológico ou estético que o texto produziu no leitor. Tipo de pergunta subdividida em duas: múltipla escolha. *Baseado no texto como você avalia o serviço florestal dos EUA? Por quê?*

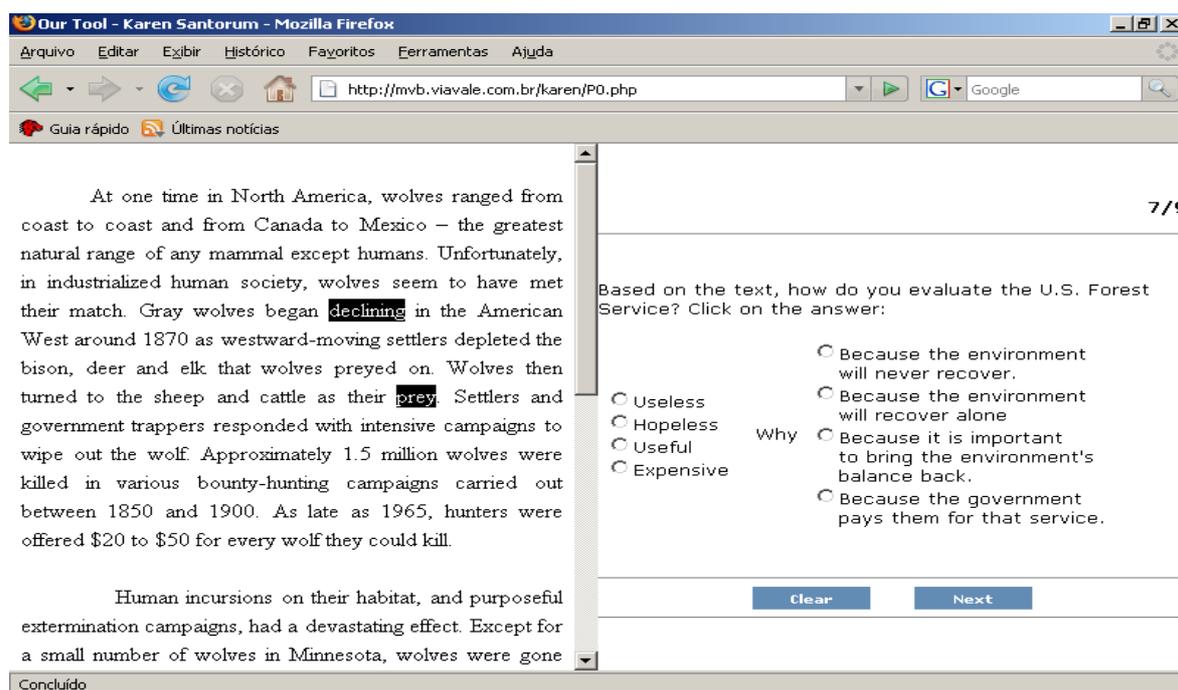



Figura 17 – Tela OUR TOOL questão 7

Questão 9: Nível taxonômico: compreensão inferencial: inferência de seqüências. Nível taxonômico: avaliação, tenta avaliar o impacto psicológico ou estético que o texto produziu no leitor. Tipo de pergunta: múltipla escolha com uma caixa de texto para o candidato digitar sua justificativa para a alternativa escolhida.

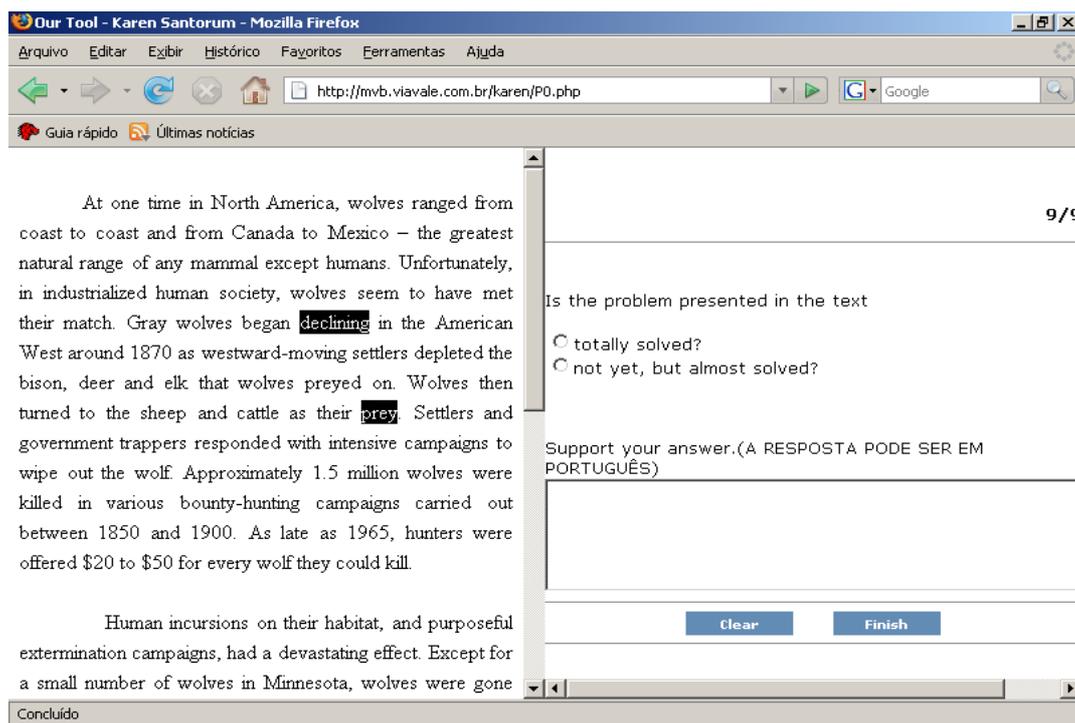



Figura 18 – Tela OUR TOOL questão 9

Depois de construído o instrumento para a pesquisa, o passo seguinte foi sujeitá-lo à etapa de validação, que subdividiu-se em validação de conteúdo e de aplicabilidade, que serão apresentadas posteriormente na seção 2.4.3.

2.4.2.3 Reconto

O terceiro instrumento utilizado na produção de dados foi o reconto do texto realizado pelos leitores após a aplicação de cada um dos testes. Segundo Brandão e Spinillo (1998), a tarefa do reconto é uma das formas de avaliação que envolve a compreensão global do texto e a seleção de aspectos relevantes do texto original e dos fatos armazenados na memória.

Como já referido na literatura em Marcuschi (1989); Brandão e Spinillo (1998); Brandão e Spinillo 2001 entre outros, as pesquisas que investigam a compreensão em leitura caracterizam-se pelo uso de um dos dois tipos de recursos metodológicos: (a) reprodução (oral ou escrita) de um texto apresentado (lido ou ouvido); ou (b) respostas a perguntas sobre um texto lido/ouvido. Poucas são as pesquisas que combinam ambas as metodologias em uma mesma amostra, como é o caso do estudo realizado por Marcuschi (1989) com adultos e adolescentes; e o estudo de Brandão e Spinillo (1998) com crianças.

A grande maioria dos estudos sobre compreensão de textos adota a segunda alternativa, procurando examinar os fatores responsáveis pela compreensão de textos, tais como: fatores lingüísticos (sintáticos, semânticos, léxicos, habilidades de decodificação); e fatores cognitivos (memória de trabalho, capacidade de monitoramento, inferências e capacidade de integrar as informações veiculadas no texto com seu conhecimento prévio) (BRANDÃO e SPINILLO, 2001).

No presente estudo, após realizar cada um dos dois testes, cada participante foi solicitado a fazer o reconto do texto lido, colocando o máximo possível de detalhes. A opção por solicitar o reconto após a realização do teste foi justamente para que se pudesse verificar se a forma com que o instrumento abordou o texto, através de suas questões, repercutiu na qualidade do reconto.

Foi considerado o fato de que o reconto por escrito de textos sofre influência da memória e da habilidade de produção de um novo texto, que caracterizam processos não específicos de leitura textual. Porém, como no reconto oral também não se teria isenção total de influências, foi mantida a opção pelo reconto escrito. Além disto, o reconto configurou-se como uma alternativa para que se desse conta do modelo cognitivo de compreensão de leitura textual aqui escolhido: o modelo de Kintsch e van Dijk referido na seção 1.3 do capítulo anterior.

Na tentativa de controlar, ao menos parcialmente, as influências do reconto por escrito, pensou-se em uma forma alternativa para que esse reconto fosse feito. Para descaracterizar um pouco a formalidade da escrita e tirar do participante a preocupação com a produção, uma vez que esse não era o foco da pesquisa, recorreu-se a uma sala de bate-papo do ambiente de Educação A Distância (EAD) da universidade. A título de ilustração,

serão reproduzidas ao longo do texto algumas telas com a produção dos participantes na sala de bate papo.

O ambiente EAD UNISC, de apoio ao ensino presencial e de suporte à educação a distância, proporciona a alunos e professores uma série de funcionalidades que visam a facilitar e otimizar as tarefas do dia-a-dia. Para apresentação e disponibilização de conteúdos, o ambiente oferece: plano de aulas, material complementar, dinâmica, eventos, novidades, glossário, bibliografia, perguntas freqüentes e gerenciador de arquivos. Quanto à comunicação, buscando incentivar a interação e cooperação entre os usuários, oferece as opções: mural, correio eletrônico, bate-papo (*chat*), fórum e mensagens on-line. Quanto à avaliação, oferece as opções: avaliação externa, avaliação *web*, tarefas e controle de notas. Quanto às ferramentas de acompanhamento das atividades do aluno pelo ambiente, há as opções: acompanhamento de tarefas e avaliação *web*, além do relatório de acessos e relatório de comunicação (www.ead.unisc.br).

Na sala de bate-papo, como o nome já diz, o participante sente-se como se estivesse "conversando" com alguém, empregando na escrita uma linguagem tipicamente oral, como podemos ver através deste exemplo:

Figura 19 – Tela bate-papo no ambiente EAD UNISC

A opção por esse suporte se deu por ter sido considerado mais prático, uma vez que o participante após realizar o teste poderia permanecer na frente do computador para fazer o reconto sem precisar de nenhum outro recurso material. A opção Bate-Papo no sistema EAD disponibiliza uma sala para conversa em tempo real, semelhante às encontradas na *Internet*, tendo, entretanto, seu acesso liberado somente no período programado para a sessão, em que as conversas podem ser registradas (gravadas) e disponibilizadas no ambiente para futura consulta, o que permitiu o acesso aos recontos para análise.

Após entrarem na sala de bate-papo, os participantes receberam uma mensagem com as instruções para o reconto, como pode ser visto na Figura 20 a seguir[44]:

Figura 20 – Tela com *e-mail* no ambiente EAD UNISC

Após realizar o reconto, ainda na sala de bate-papo, cada participante respondeu ao todo a quinze questões (sete após o primeiro teste e oito após o segundo teste) sobre seu

[44] Embora cientes dos erros de concordância verbal no texto do e-mail, optou-se por uma linguagem informal, condizente com a que é utilizada em ambientes de bate-papo.

perfil (trajetória na aprendizagem do inglês) e sobre suas estratégias de leitura, além de compararem as duas ferramentas. Esses dados, que correspondem à parte da análise qualitativa da pesquisa, também foram tabulados em uma planilha para posterior análise. Uma vez definidos os instrumentos de produção de dados, partiu-se para a validação dos mesmos.

2.4.3 Validação do instrumento: conteúdo e aplicabilidade

A validação de conteúdo, conforme o *Dictionary of Language Testing* (2002), requer o julgamento de examinadores especialistas para avaliar a adequação e a representatividade dos itens do instrumento. Para consolidar esse tipo de validação, o texto e as questões da OUR TOOL foram enviados por correio eletrônico, em formato de texto padrão do *word*, para três juízes, pedindo-lhes que respondessem ao teste – para verificação do gabarito e a pertinência das questões. Foi explicado que se tratava de uma proposta alternativa de avaliação da compreensão leitora, porém sem maiores detalhes e sem mencionar as dimensões cognitivas contempladas por cada questão para não influenciar na observação. A única informação dada foi que posteriormente o instrumento seria realizado em uma versão *on line,* por isso o verbo “clicar” nas instruções das questões.

A escolha pelos examinadores se justifica pela experiência dos mesmos com o ensino de língua inglesa. O primeiro deles é mestre, com mais de dez anos de experiência em docência no ensino universitário, com ênfase em leitura em L2. O segundo juiz possui experiência de trabalho com a língua inglesa em diferentes séries do ensino fundamental e médio. O terceiro juiz tem experiência de ensino em cursos particulares de inglês em diversos níveis.

Os juízes responderam também por correio eletrônico enviando o teste respondido, com suas observações. Em alguns casos, essas observações foram discutidas *on line* em um programa de mensagens instantâneas.

O primeiro juiz considerou o instrumento claro e as questões bem elaboradas, porém fez algumas observações. Através de seu olhar, foram constatados pequenos problemas nos enunciados das questões que poderiam atrapalhar o candidato. Na versão enviada a esse juiz

aparecia como instrução, antes do primeiro item, que as questões poderiam ser respondidas em português; no entanto, apenas na última questão havia essa possibilidade. Assim, a instrução foi retirada e colocada apenas no enunciado da última questão.

Outra sugestão, também aceita, foi em relação à questão que contempla o nível taxonômico leitura crítica, julgamento de valores. O juiz sugeriu que fosse colocado no enunciado da questão que a resposta deveria ser com base no texto.

O segundo juiz também declarou como aprovado o instrumento, porém considerou inferencial demais a sexta questão. A questão foi mantida na forma em que estava, uma vez que contempla exatamente o nível taxonômico da compreensão inferencial, inferência de causa e efeito que se pretendia avaliar.

O terceiro juiz fez observações em relação a quinta questão chamando a atenção para o fato de que ela apresentava mais de uma opção de resposta. A questão foi mantida da forma em que estava por contemplar o nível taxonômico da compreensão inferencial, inferência de detalhes que busca verificar a capacidade do leitor de conjecturar sobre os detalhes adicionais que o autor poderia ter incluído no texto tendo, intencionalmente, mais de uma alternativa correta.

Apesar das observações feitas pelos juízes, o instrumento foi igualmente aprovado pelos três e as questões julgadas bem elaboradas e pertinentes com níveis de leiturabilidade e legibilidade apropriados para o nível de proficiência da amostra. Após a validação de conteúdo, foi formatada uma nova versão do instrumento e feitas as devidas modificações nas instruções conforme as sugestões acatadas. Assim, foi dado prosseguimento à próxima etapa da validação: a validação da aplicabilidade do instrumento.

A segunda etapa da validação, ou seja, a validação de aplicabilidade visa avaliar a adequação do instrumento quanto ao suporte escolhido e quanto ao seu funcionamento no momento da realização pelos candidatos. Para consolidar esse tipo de validação foi executado um teste piloto com um colaborador que não participou da etapa posterior de produção de dados. Nesse estudo piloto, optou-se por rodar todo o procedimento: os dois testes (TOEFL e OUR TOOL), seguidos das etapas de reconto e investigação sobre

estratégias de leitura, ao invés de testar apenas a aplicabilidade da ferramenta desenvolvida para o estudo, a fim de se ter a noção de tempo e de como seria a testagem.

Assim, os objetivos com o estudo piloto foram os de 1) investigar a adequação do instrumento construído para a pesquisa; 2) verificar o tempo de aplicação do mesmo, bem como o tempo total de duração das sessões. Finalizada a etapa de validação partiu-se para a aplicação dos instrumentos.

2.4.4 Aplicação dos instrumentos e procedimentos específicos

Os vinte participantes foram divididos randomicamente em dois grupos de dez e passaram igualmente por todas as etapas da produção de dados, porém em diferentes seqüências. O primeiro grupo, que continha os participantes de 1 a 10, seguiu a seqüência 1 que corresponde à realização primeiramente do TOEFL e depois da OUR TOOL. O segundo grupo, com sujeitos de 11 a 20 , seguiu a seqüência 2, correspondente à realização da OUR TOOL em primeiro lugar e depois o TOEFL. Após cada teste os participantes realizaram o reconto e responderam às questões sobre perfil e estratégias de leitura na sala de bate-papo.

No intuito de facilitar o acompanhamento das etapas de produção de dados, foi desenvolvido o seguinte quadro com as duas seqüências aplicadas:

<table>
<tr><th>ETAPAS</th><th>GRUPO 1</th><th>GRUPO 2</th></tr>
<tr><td>1</td><td>TOEFL</td><td>OUR TOOL</td></tr>
<tr><td>2
sala de bate-papo</td><td colspan="2">Reconto do texto</td></tr>
<tr><td>3
sala de bate-papo</td><td colspan="2">Perfil e estratégias de leitura</td></tr>
<tr><td>4</td><td>OUR TOOL</td><td>TOEFL</td></tr>
<tr><td>5
sala de bate-papo</td><td colspan="2">Reconto do texto</td></tr>
<tr><td>6</td><td colspan="2">Estratégias de leitura</td></tr>
</table>

sala de bate-papo	

Quadro 6 – Etapas da produção de dados

As duas seqüências permitiram o cruzamento dos dados de modo a analisar a performance em cada abordagem independente da seqüência e evitar alterações no reconto por questões de memória.

Os dados foram produzidos em várias etapas, algumas delas individualmente, de acordo com a disponibilidade dos participantes e possibilidade de uso do laboratório de informática, pois o mesmo tinha que ser reservado e ficava fechado durante a realização do experimento. Em algumas dessas etapas foi possível reunir alguns participantes, o que otimizou o tempo. Nas sessões em que o grupo era maior, o monitor do laboratório, que já estava ciente das etapas, auxiliou no processo de salvar e arquivar as páginas de resultado dos testes dos participantes.

Tendo sempre em vista o desenvolvimento da compreensão leitora e para que a pesquisa pudesse dar um retorno mais imediato para a comunidade, ao final de cada sessão de produção de dados, foi proposta uma conversa com os participantes interessados acerca de seus erros e das dificuldades encontradas. Foram enviadas, por correio eletrônico, para cada um, as páginas de respostas dos dois testes e o *link* com o endereço eletrônico de onde está hospedada OUR TOOL na *internet* para que, quem quisesse, pudesse fazer o teste novamente como exercício, ou reler o texto, ou ainda, rever as questões.

Outra providência tomada foi enviar à professora do CELINC, que estava ciente dos objetivos da pesquisa, os resultados e a ferramenta para que ela também pudesse discutir com seus alunos em aula, fazendo da experiência uma prática pedagógica. Finalizada essa etapa, partimos para a discussão e análise dos dados que serão descritas no próximo capítulo.

3 APRESENTAÇÃO DOS DADOS E DISCUSSÃO DOS RESULTADOS OBTIDOS A PARTIR DA LENTE METODOLÓGICA ADOTADA

> A avaliação de leitura deveria incluir amostras sistemáticas de comportamentos de leitura que guiem as decisões acerca da habilidade de leitura (JOHNSTON, 1983)[45].

A apresentação dos dados deste estudo será feita em etapas, assim como se deu a produção dos mesmos. Ao final de cada etapa ocorrerá a discussão dos resultados apresentados. Na primeira etapa serão contemplados os resultados obtidos pelos participantes do grupo 1 e 2 nos dois testes, bem como a comparação entre os mesmos. Na segunda etapa serão apresentados os resultados da análise e classificação dos recontos. Nessa etapa serão comparadas as classificações dos recontos realizados pós TOEFL e pós OUR TOOL. Na terceira etapa será verificada a correlação entre os resultados dos testes e as classificações dos recontos. A última etapa, que contempla a análise das respostas sobre as estratégias de leitura dos participantes e seus respectivos comentários sobre os testes, não ocorrerá em um momento separado. Esses resultados serão apresentados ao longo desse capítulo, simultaneamente às etapas anteriores, a fim de iluminar a discussão.

No presente estudo, a avaliação da compreensão, a partir dos resultados da pesquisa, é observada através de lentes específicas, como referido no capítulo anterior. São utilizadas as lentes do modelo de Kintsch e van Dijk (1978) à luz do critério de autenticidade pertencente ao modelo de utilidade de Bachman e Palmer (1986), apresentados no capítulo 2. Porém, em momento algum se pretende aqui apontar uma forma de avaliação como sendo "a correta" em detrimento de outra, uma vez que acreditar na existência de uma medida avaliativa como sendo a melhor para qualquer situação dada seria o mesmo que estar às cegas.

3.1 Resultados dos testes de avaliação da compreensão

Nessa primeira etapa da análise, foi observado o desempenho dos participantes nos dois testes. As Tabelas 1 e 2 a seguir mostram a proporção de acertos no TOEFL e no OUR TOOL, respectivamente, dentro de cada seqüência. Nas tabelas, os participantes estão divididos em dois grupos de dez, de acordo com a seqüência de realização das provas. Conforme a Tabela 1, 20% dos sujeitos do grupo 1 (TOEFL ⇨ OUR TOOL) acertaram 50% da prova; 10% acertaram 66,67%; 20% acertaram 83,33% e 50% acertaram 100% da prova.

[45] No original: *Reading assessment should include systematic samplings of reading behaviors that guide decisions about reading ability* (JOHNSTON, 1983 apud FARR e CAREY 1986, p. 117).

No grupo 2 (OUR TOOL TOEFL) 10% da amostra acertaram 50% da prova; 40% acertaram 83,33 e 50% acertaram 100% da prova.

Tabela 1 – Proporção de acertos no TOEFL

Proporção de acertos no TOEFL	**GRUPO 1**		**GRUPO 2**		**Média entre os grupos**	
	n	**%**	**n**	**%**	**n**	**%**
50,00	2	20,0	1	10,0	3	15,0
66,67	1	10,0	0	0,0	1	5,0
83,33	2	20,0	4	40,0	6	30,0
100,00	5	50,0	5	50,0	10	50,0
Total	**10**	**100,0**	**10**	**100,0**	**20**	**100,0**

Tabela 2 – Proporção de acertos no OUR TOOL

Proporção de acertos no OUR TOOL	**GRUPO 1**		**GRUPO 2**		**Média entre os grupos**	
	n	**%**	**n**	**%**	**n**	**%**
47,37	2	20,0	0	0,0	2	10,0
52,63	2	20,0	1	10,0	3	15,0
57,89	0	0,0	2	20,0	2	10,0
63,16	0	0,0	1	10,0	1	5,0
68,42	1	10,0	2	20,0	3	15,0
73,68	3	30,0	3	30,0	6	30,0
84,21	1	10,0	1	10,0	2	10,0
89,47	1	10,0	0	0,0	1	5,0
Total	**10**	**100,0**	**10**	**100,0**	**20**	**100,0**

Na Tabela 2, verifica-se que 20% dos sujeitos do grupo 1 acertaram 47,37% da OUR TOOL; 20% acertaram 52,63%; 10% acertaram 68,42%; 30% acertaram 73,68%; 10% acertaram 84,21% e 10% acertaram 89,47%. Entre os sujeitos do grupo 2, vê-se que 10% da amostra acertaram 52,63%; 20% acertaram 57,89%; 10% acertaram 63,16%; 20% acertaram 68,42%; 30% acertaram 73,68% e 10% acertaram 84,21%. Para facilitar a visualização, é

apresentada a Figura 21 a seguir, em forma de gráfico, com o número de casos e a proporção de acertos em ambos os testes.

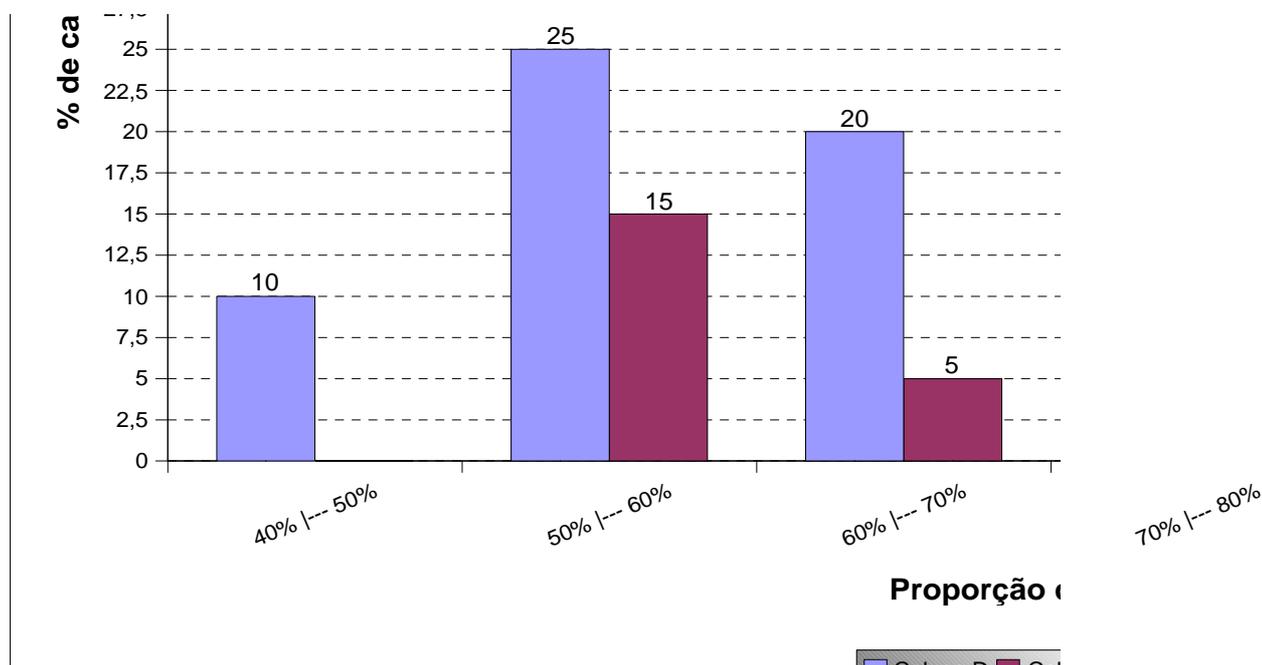



Figura 21 – Proporção de acertos nos testes TOEFL e OUR TOOL.

Para verificar se a seqüência de realização dos testes influenciou no resultado, foi aplicado o teste t de Student, que verifica a existência de diferenças estatisticamente significantes entre as médias de dois grupos independentes. As médias e desvios padrão são apresentados na Tabela 3 a seguir:

Tabela 3 – Teste t de Student para amostras independentes

	Seqüência de realização dos testes	**n**	**Média**	**Desvio-Padrão**
Proporção de acertos no TOEFL (6 questões)	TOEFL ⇨ OUR TOOL	10	83,33	20,78699
	OUR TOOL ⇨ TOEFL	10	88,33	15,81139
Proporção de acertos no OUR TOOL (19 questões)	TOEFL ⇨ OUR TOOL	10	66,31	15,33458
	OUR TOOL ⇨ TOEFL	10	67,37	9,54489

Conforme conclusão do teste, não há diferença estatisticamente significativa entre as proporções médias de acertos no TOEFL do grupo que fez esse teste em primeiro lugar (média de 83 %) em comparação com o grupo 2 que fez o teste em segundo lugar (média

88%), sendo p = 0,55. Assim como não há diferença estatisticamente significativa entre as proporções médias de acertos no teste OUR TOOL do grupo 1 (média de 66%), que fez primeiro o TOEFL, em comparação com o grupo 2 (média de 67%) que fez primeiro o OUR TOOL (p = 0,85).

Concluído que a seqüência de realização dos testes não interferiu no resultado, a próxima verificação feita foi em relação à diferença entre as médias de acertos nos testes dos dois grupos. Para isso foi rodado o teste t de Student para amostras emparelhadas, cujo resultado é apresentado na Tabela 4 a seguir.

Tabela 4 – Teste t de Student para amostras emparelhadas

	Média	**n**	**Desvio-Padrão**
Proporção de acertos no TOEFL (6 questões)	85,8333	20	18,15705
Proporção de acertos no OUR TOOL (19 questões)	66,8421	20	12,44319

O teste concluiu que há diferença estatisticamente significativa entre a proporção média de acertos no TOEFL em comparação com a proporção média de acertos no OUR TOOL ($p < 0,001$). Conforme a Tabela 4, o índice médio de acertos no TOEFL foi de 85%, maior que no OUR TOOL, que teve como média 66%. Como mostrado anteriormente na Figura 16, não houve 100% de acerto no OUR TOOL nem mesmo quando esse foi realizado após o TOEFL, situação essa em que os participantes estavam lidando com o mesmo texto pela segunda vez, depois de já terem respondido sobre o mesmo. Já no TOEFL, 50% da amostra gabaritou a prova.

Uma possível explicação para essa diferença no índice de acertos entre as duas provas, e que será retomada posteriormente, pode estar na abordagem ao texto feita por cada instrumento. A abordagem no OUR TOOL, que contemplou as dimensões cognitivas e afetivas propostas pela taxonomia de Barret (1968), mostrou-se mais difícil quando comparada com a abordagem do TOEFL, não apenas nos índices de acerto mas também nos comentários feitos pelos leitores na sala de bate-papo após a realização das provas e dos recontos.

Nessa sala foi solicitado aos participantes que respondessem se as questões do teste feito dependiam diretamente do texto para serem respondidas, ou seja, se tinham suas

respostas imediatamente localizáveis no texto. Os resultados, que podem ser vistos a partir da Tabela 5 a seguir, mostram que a grande maioria (95%) dos participantes apontou as questões do TOEFL como sendo mais diretamente reconhecíveis no texto e, por isso, mais dependentes do texto para serem respondidas, enquanto no OUR TOOL apenas 45% dos sujeitos apontou tal característica.

Tabela 5 – Questões diretamente respondidas a partir do texto

		n	**%**
As questões do teste dependiam diretamente do texto para serem respondidas? (TOEFL)	Sim	19	95,0
	Não	1	5,0
	Total	**20**	**100,0**
As questões do teste dependiam diretamente do texto para serem respondidas? (OUR TOOL)	Sim	9	45,0
	Não	11	55,0
	Total	**20**	**100,0**

Como cada sujeito respondeu duas vezes a mesma questão – uma pós TOEFL e outra pós OUR TOOL, foi rodado o teste de McNemar para verificar se houve mudança significativa nas respostas entre uma situação e outra. Nesse caso, a conclusão do teste foi de que houve mudanças estatisticamente significativas entre as respostas dadas pelos sujeitos para as perguntas relativas aos testes TOEFL e OUR TOOL ($p = 0,002$). A fim de reforçar o dado quantitativo, serão apresentadas aqui algumas das respostas dadas a essa pergunta na sala de bate-papo, primeiramente em relação ao OUR TOOL e depois ao TOEFL, correspondendo ao dado qualitativo da pesquisa que contribuiu nas reflexões.

Figura 22 – Tela bate-papo pós OUR TOOL - questão 3 - seqüência: OUR TOOL - TOEFL

Uma leitura possível desses comentários é a de que as questões no teste OUR TOOL não viabilizavam a atitude de “copiar e colar” para serem respondidas. Precisavam de uma interpretação do texto como um todo. A resposta na Figura 23 refere-se ao TOEFL.

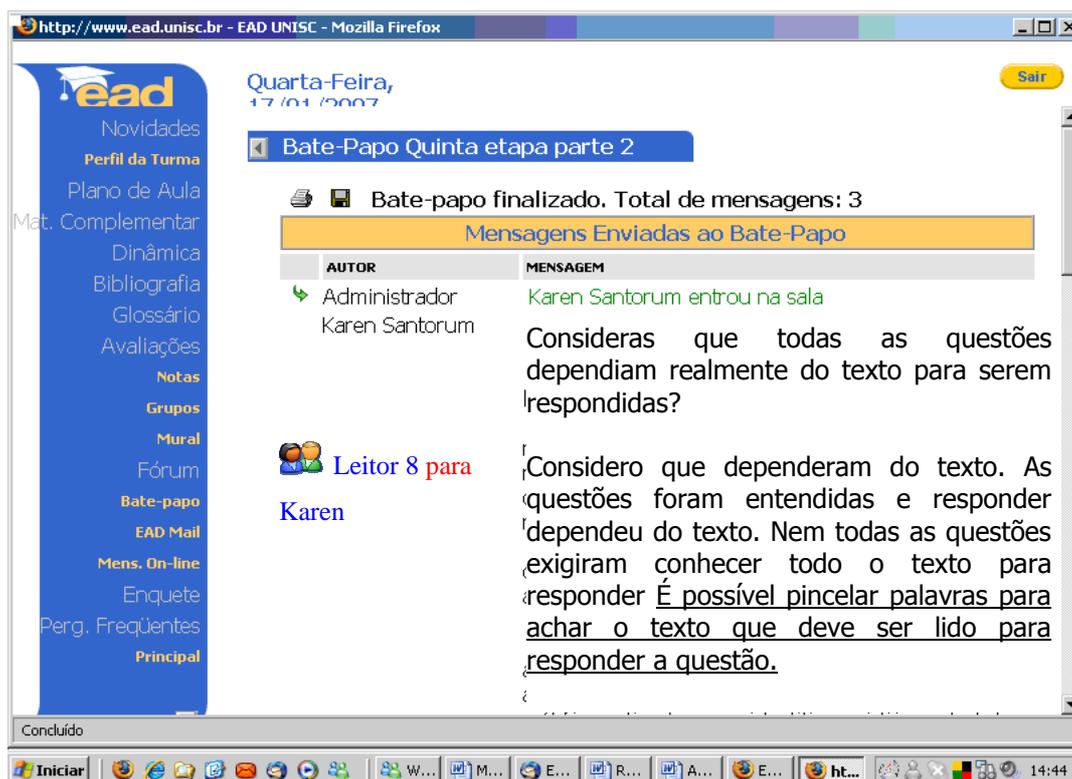



Figura 23 – Tela bate-papo pós TOEFL - questão 3 - seqüência: TOEFL – OUR TOOL

A partir dessa resposta é possível identificar uma estratégia de leitura aplicada por um leitor, já atestado como proficiente em língua inglesa, muito necessária em tarefas de avaliação mas não suficientes em tarefas de vida real quando há necessidade de compreensão. A estratégia usada informa acerca da habilidade desse leitor em lidar com o texto mas não necessariamente sobre sua capacidade de compreensão. Esse participante obteve resultado de 100 % no TOEFL e 47 % no OUR TOOL.

À medida que a questão utiliza um vocabulário similar ao do texto, ela pode estar permitindo que uma estratégia como essa seja utilizada, o que pode vir a alterar o resultado. No caso do TOEFL, conforme relato de Peirce (1992), essa era uma tendência entre os revisores da equipe de desenvolvimento da prova: utilizar um vocabulário semelhante, bem como a ordenação das questões seguindo a ordem do texto, conforme observado no seguinte relato: “Além disso, eu considerei que seria em geral útil aos candidatos se a ordem dos itens no conjunto seguisse a ordem das informações do próprio texto”[46].

[46] No original: *Furthermore, I assumed it would be generally helpful to candidates if the order of items in the set followed the order of information in the text itself* (PEIRCE, 1992, p. 671).

Em outra resposta, o leitor indica uma diferença de comportamento de leitura em cada teste em função da abordagem. Essa diferença remete àquela oposição estabelecida por Samuels e Kamil (1988), já referida no capítulo 2, em que os autores afirmam que a compreensão não é o processo de *retirar* o sentido de um texto, mas sim o de *trazer* sentido ao mesmo. Nos comentários a seguir os leitores "confessam" o procedimento de retirada do texto:

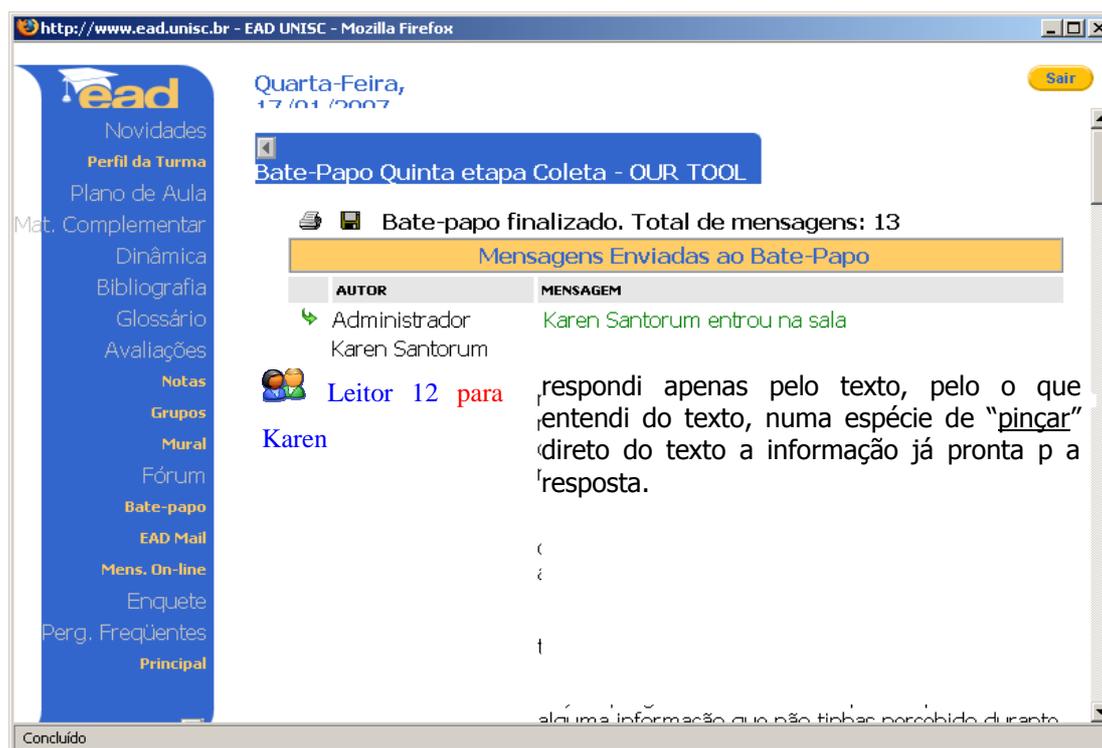



Figura 24 – Tela bate-papo pós TOEFL – questão 2 – seqüência: OUR TOOL - TOEFL

Em outras respostas relativas ao TOEFL, apresentadas na Figura 25 a seguir, o mesmo processo de extração é identificado em uma estratégia de leitura exploratória, conforme apresentado no quadro 2 no referencial teórico, utilizada pelos leitores:

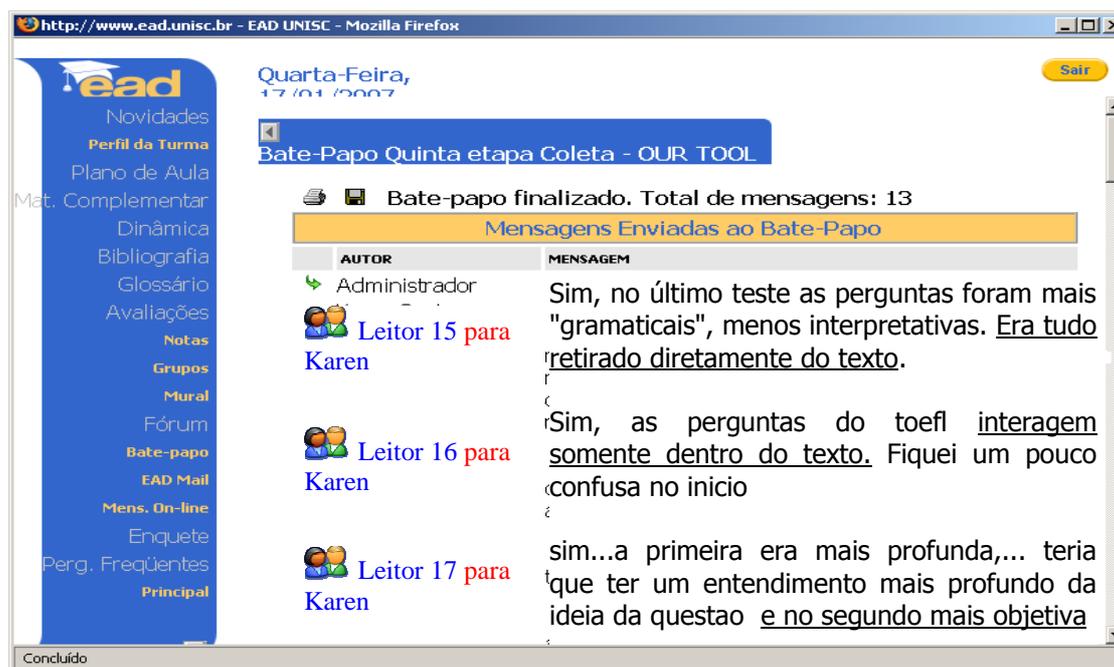



Figura 25 – Tela bate-papo pós TOEFL – questão 7 – seqüência: OUR TOOL - TOEFL

A estratégia de "retirada", quando bem dominada pelos candidatos, confere-lhes ótimos resultados no teste TOEFL. A saber, o leitor 15 do comentário anterior obteve resultado de 100 % no TOEFL e 68 % no OUR TOOL; o leitor 16 obteve 83 % no TOEFL e 63 % no OUR TOOL e o leitor 17 obteve 83 % no TOEFL e 52 % no OUR TOOL. Todos aprovados na seção simulada de leitura do TOEFL.

A diferença na abordagem ao texto através das questões de ambos os testes vem novamente à tona quando os participantes são solicitados a compararem as questões das duas provas. Conforme mostrado na Tabela 6 a seguir, uma média de 90% dos participantes (100% no grupo 1 e 80% no grupo 2) apontaram a abordagem do OUR TOOL como sendo a que exigiu maior profundidade na compreensão (gerando um modelo situacional mais completo), em contraste com o comportamento de retirada da superfície do texto (base textual) apresentado no TOEFL.

Tabela 6 – Diferença de abordagem

Em qual dos 2 testes as perguntas fizeram ir ao texto com mais profundidade?	Seqüência de realização dos testes					
	Grupo 1 TOEFL ▶ OUR TOOL		Grupo 2 OUR TOOL ▶ TOEFL		Média entre os grupos	
	n	%	n	%	n	%
OUR TOOL	10	100,0	8	80,0	18	90,0
TOEFL	0	,0	1	10,0	1	5,0
Indiferente	0	,0	1	10,0	1	5,0
Total	10	100,0	10	100,0	20	100,0

Para fins de análise estatística as respostas foram codificadas, mas a forma original como foi feita a pergunta, na sala de bate-papo do ambiente EAD, elicitou comentários dos participantes que auxiliaram a reflexão. A Figura 26 a seguir apresenta esses comentários feitos por participantes do Grupo 1, que realizou primeiro o TOEFL e depois OUR TOOL:

Figura 26 – Tela bate-papo pós OUR TOOL – questão 7 – seqüência: TOEFL – OUR TOOL

O leitor 6 aponta nesse comentário a necessidade de uma compreensão total desde as perguntas até o texto em si na OUR TOOL. Além disso ressalta o fato de que as perguntas nesse instrumento não remetem diretamente a uma parte do texto - o que facilitaria a localização da resposta - exigindo a construção e compreensão do sentido. Esse participante obteve resultado de 100% no TOEFL e 73% no OUR TOOL.

O leitor 4 também aponta a diferença entre as abordagens e um maior nível de exigência no OUR TOOL, indicando a necessidade de mais "dicionário". Talvez esse leitor tenha querido indicar necessidade de maior vocabulário para ter uma maior compreensão do texto para poder responder às questões do teste. Esse participante apresentou os mesmos escores que o primeiro: 100 % no TOEFL e 73 % no OUR TOOL.

Interessante foi a constatação de que embora o índice de acerto no OUR TOOL tenha sido consideravelmente menor do que no TOEFL no geral da amostra, o nível de representação do texto, verificado nos recontos, foi maior exatamente na ferramenta em que os participantes tiveram mais dificuldades para responder e pontuaram menos. Esses dados serão apresentados na segunda etapa da análise a seguir.

3.2 Resultados da análise e classificação dos recontos

Para análise dos recontos foi utilizado o modelo de Kintsch e van Dijk (1978), referido no capítulo 2. No modelo, distinguem-se três níveis na representação do discurso: 1) a estrutura de superfície (palavras e sintagmas do texto); 2) o conteúdo semântico local (microestrutura) e global (macroestrutura) e 3) o modelo de situação. O modelo de situação, no qual a situação descrita no texto é representada, é como um modelo mental, isto é, um construto localizado na memória episódica que contém a representação subjetiva do texto, incluindo inferências e experiências prévias (KINTSCH, 1998).

Assim, com base na observação do modelo situacional, os recontos foram analisados e enquadrados em categorias, observando: a) a recuperação das idéias centrais do texto, e b) a quantidade de inferências, interferências e reconstruções presentes.

A classificação em inferências, interferências e reconstruções seguiu parâmetros propostos por Parente, Saboskinski, Ferreira e Nespoulous (1999). Inferências são os comentários pertinentes a respeito de fatos presentes no texto original ou o relato de um fato que deve ter ocorrido, mas que está ausente no texto original. Interferências ocorrem quando o sujeito modifica o significado das cláusulas do texto por misturar elementos presentes, porém independentes. Reconstrução é a introdução de cláusulas que relatam fatos não presentes no texto original e que não poderiam ser inferidos pelas informações contidas no texto (SALLES, 2005).

As categorias propostas para análise dos recontos, apresentadas no Quadro 7 a seguir, foram elaboradas com base em categorias pré-estabelecidas existentes na literatura sugeridas por Brandão e Spinillo (2001). A decisão pela não utilização das categorias já existentes e sim pela elaboração de novas ocorreu em função de que algumas dessas categorias acabavam por ficar amplas demais e, por isso, não davam conta de demonstrar a diferença de qualidade do modelo situacional entre os recontos. De um modo geral, vários recontos se encaixavam em uma mesma categoria, porém havia grandes diferenças entre eles que não eram contempladas pela mesma. Para efeito de análise, o texto foi dividido em dez orações principais (ANEXO D).

CATEGORIAS	DESCRIÇÃO
I	Reproduções desconectadas do texto, ou recontos que limitam-se a frases que marcam a abertura e o fechamento do texto. Reproduzidas aqui essencialmente as orações 2 e 8.
II	Reproduções que se limitam a alguns eventos presentes em determinados blocos do texto original. Reconto feito de forma desarticulada, não apresentando nenhuma relação causal e incluindo informações pouco vagas em relação ao texto original. Ocorrem interferências. Reproduzidas aqui essencialmente as orações 2, 6 e 8.
III	Reproduções que, embora apresentem cadeias causais e envolvam alguns personagens e alguns eventos presentes em determinados blocos do texto original, não trazem alguns detalhes importantes para a representação do texto, como a oração 6, por exemplo. Ocorrem inferências. Reproduzidas aqui essencialmente as orações 2, 3, 4, 5 e 8.
IV	Reproduções que apresentam a cadeia causal somada a detalhes pertinentes do texto original. Embora não haja uma referência explícita ao desfecho do texto, ocorre uma clara percepção da quebra de paradigma presente no texto. Ocorrem inferências. Reproduzidas aqui essencialmente as orações 1, 2, 3, 4, 5, 6, 7 e 8.

V	Reprodução completa, em que as idéias do texto são integralmente reproduzidas. Surgem inferências mais elaboradas que ocorrem de forma bastante articulada. O reconto segue um eixo onde o problema central é apresentado e resolvido, apontando os meios para tal. As relações causais são apresentadas de forma clara.

Quadro 7 – Classificação dos recontos em categorias

A partir das novas categorias foi feita a classificação. Os recontos foram agrupados randomicamente, sem identificação de grupo, seqüência ou sujeito, a fim de evitar um olhar tendencioso no momento da classificação em categorias. Posteriormente, para fins de comparação, esses recontos foram identificados e contabilizados. Os resultados são mostrados na tabela 7 a seguir.

Nessa tabela, nas categorias pós TOEFL, é mostrado que a maioria dos leitores (80%) elaborou recontos que ficaram nas categorias mais baixas (I e II). Nessas categorias foram citados alguns eventos do texto, porém de forma desarticulada e sem apresentar a cadeia causal. Apenas 5% da amostra apresentou reconto enquadrado na categoria IV e 5% na categoria V. Já em relação às categorias pós OUR TOOL, a Tabela 7 mostra que 60% dos recontos foram classificados nas categorias I e II; 15% na categoria IV e 15% na categoria V.

Tabela 7 – Categorias de reconto pós TOEFL e pós OUR TOOL

	Pós-TOEFL		**Pós- OUR TOOL**	
Categorias	**n**	**%**	**n**	**%**
I	9	45,0	5	25,0
II	7	35,0	7	35,0
III	2	10,0	2	10,0
IV	1	5,0	3	15,0
V	1	5,0	3	15,0
Total	**20**	**100,0**	**20**	**100,0**

Os leitores que tiveram seus recontos enquadrados nas duas primeiras categorias apresentam dificuldades na compreensão de leitura textual. Recontos inseridos nessas categorias apresentam, além da ausência de cadeia causal, muitas reconstruções, ou seja

inserções de fatos e idéias que não estavam presentes no texto original, e que não poderiam com as informações disponíveis, serem seguramente inferidos. Um exemplo de reconstrução aparece grifado no seguinte reconto:

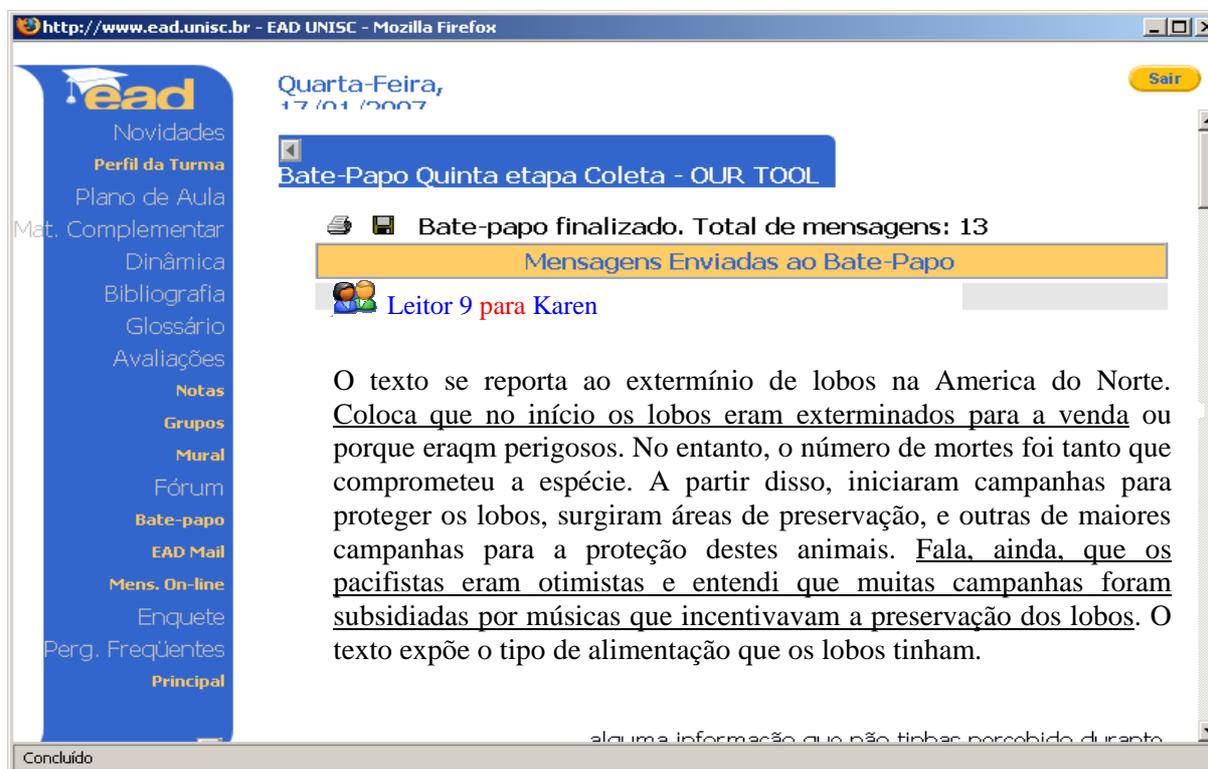



Figura 27 – Tela bate-papo Reconto pós TOEFL – Seqüência TOEFL – OUR TOOL

Todavia, conforme apontado por Salles (2005), cabe pensar até que ponto essas reconstruções evidenciam uma menor compreensão do texto lido ou servem como uma estratégia para preencher as lacunas da compreensão.

As dificuldades em estabelecer as cadeias causais encontradas em 80% dos recontos pós TOEFL e em 60% pós OUR TOOL, que ficaram nas categorias I e II, podem indicar que a compreensão de orações é atingida – tanto que conseguiram responder as perguntas dos testes – porém há dificuldade de integrar essas orações em um todo coerente. Conforme afirma Marotto (2000), citado por Salles (2005), a compreensão do texto não é a soma do significado de suas orações individuais. O leitor pode não chegar a estabelecer as relações de coerência que dão sentido ao texto, mesmo entendendo as orações individuais.

Ainda conforme Salles (2005), os modelos mentais elaborados durante a compreensão têm certamente elementos que são externos à história, relacionados à bagagem individual do leitor. Como os esquemas e os modelos mentais são individuais, criados a partir da vivência e conhecimento prévio de cada um, e como os fatos são lembrados por meio desses esquemas individuais, o texto evocado nunca é exatamente igual ao texto lido (SALLES e PARENTE, 2004). Segundo Leffa (1996), isso ocorre porque, já no ato de decodificação, o leitor está contribuindo com a sua experiência para a percepção do que lê, além de o lapso de tempo entre a compreensão e a evocação também contribuir para modificar o texto de acordo com os esquemas individuais.

Os recontos enquadrados na categoria III apresentam alguns blocos do texto, inserindo o problema e sua resolução, porém não trazem alguns detalhes importantes responsáveis por garantir uma relação causal bem sustentada. As categorias IV e V conforme descritas no Quadro 8, são as que indicam uma maior representação do texto por serem as mais completas e apresentarem uma cadeia causal bem estabelecida em que as idéias centrais do texto são reproduzidas. Nessas categorias ocorrem inferências e estão ausentes as reconstruções ou interferências.

A Tabela 8 a seguir mostra que em todos os casos da amostra foi constatada ou a manutenção da categoria de reconto ou o aumento da mesma pós OUR TOOL, mas jamais a diminuição dela. Conforme a linha diagonal da tabela, doze casos permaneceram na mesma categoria nos dois testes e oito (acima da linha) atingiram uma categoria mais elevada após OUR TOOL.

Tabela 8 – Categorias de reconto pós TOEFL * Categorias de reconto pós OURTOOL

Categorias de reconto pós TOEFL	**Categorias de reconto pós OURTOOL**					**Total**
	1	**2**	**3**	**4**	**5**	
1	5	2	2	0	0	**9**
2	0	5	0	2	0	**7**
3	0	0	0	0	2	**2**
4	0	0	0	1	0	**1**
5	0	0	0	0	1	**1**
Total	**5**	**7**	**2**	**3**	**3**	**20**

Para verificar se houve uma mudança substancial nas categorias de reconto correspondentes a cada um dos testes foi aplicado o teste da homogeneidade marginal. Esse teste é uma extensão do teste de McNemar para variáveis categóricas, com mais de duas alternativas de resposta. O teste da homogeneidade marginal verifica mudanças nas respostas devidas a alguma forma de intervenção experimental utilizando a distribuição chi-quadrado que verifica a existência de associação entre variáveis categóricas. O teste concluiu que houve mudança estatisticamente significativa entre as categorias de reconto relativas ao teste TOEFL e ao OUR TOOL ($p = 0,006$), sendo essa mudança para mais, como visto na Tabela 8, o que nos permite afirmar que houve uma melhora na compreensão textual pós OUR TOOL tendo em vista a produção de recontos mais ricos e fiéis ao texto alvo, conforme Quadro 8, da classificação dos recontos em categorias.

Esses resultados vêm confirmar a hipótese de que dependendo do grau de profundidade da abordagem ao texto, o nível de representação pode ser maior ou menor. A pontuação no teste OUR TOOL foi menor em função de uma abordagem mais profunda, como visto na etapa anterior a partir dos comentários dos participantes e dos índices. Contudo, a representação textual após esse teste foi maior exatamente em conseqüência dessa abordagem. A Figura 28 traz comentários dos participantes atribuindo à abordagem da OUR TOOL um maior entendimento do texto, o que pode ter garantido uma melhor classificação do modelo situacional.

Figura 28 – Tela bate-papo pós OUR TOOL – questão 4 – Seqüência TOEFL – OUR TOOL

A melhora nas categorias pós OUR TOOL poderia, à primeira vista, surpreender se o resultado (nota) obtido nos testes fosse tomado como parâmetro. Se a interpretação dos resultados fosse a de que escores mais altos (como no TOEFL) representariam maior entendimento e escores mais baixos (como no OUR TOOL) representariam um menor entendimento então, esperar-se-ia uma correlação entre os resultados e a previsão seria a de que os recontos pós TOEFL seriam de categorias mais altas.

Interessante observar que os próprios participantes, ao verem os resultados de ambos os testes, ficavam impressionados pelos mesmos e acreditavam, em um primeiro momento, ter entendido mais do texto no TOEFL do que no OUR TOOL, influenciados pelas notas

obtidas. Uma leitura possível disso é a confiança cega que as pessoas depositam nos números, sem se questionarem sobre o que de fato eles significam. É nesse sentido que uma **mudança de foco na avaliação da compreensão** se faz necessária na renovação da prática pedagógica. Caberia perguntar se um escore de 100% no TOEFL significa 100% de compreensão, ou apenas 100% de acerto num tipo de teste do qual foi possível inferir o funcionamento, ou seja, qual o truque (fornecido a partir de “treinamento”) para obter êxito. A Figura 29 a seguir mostra alguns comentários feitos pelos participantes na sala de bate-papo após a realização do OUR TOOL, indicando um descontentamento com sua performance nesse teste:

Figura 29 – Tela bate-papo pós OUR TOOL – questão 3 – seqüência: TOEFL - OUR TOOL

Outro quesito na concepção de avaliação da compreensão que mereceria ser repensado em novas práticas pedagógicas surgiu ao longo da pesquisa ainda na fase de construção da ferramenta. Na etapa da validação da OUR TOOL foi possível perceber uma amostra do efeito *washback*, ou seja, do impacto da avaliação tradicional na pedagogia

escolar. Embora a intenção da presente investigação não fosse a de coletar indícios desse impacto, mas sim investigar as possíveis conseqüências do mesmo no sistema educacional, esse dado surgiu de maneira espontânea, não podendo deixar de ser mencionado.

Não foram fornecidos aos juízes detalhes sobre a teoria subjacente ao teste. Nem lhes foi dito que se tratava de uma forma alternativa de avaliação da compreensão não tão convencional para evitar qualquer forma de indução.

Os juízes cumpriram com seu papel, avaliando o instrumento, porém foi possível perceber em alguns comentários a incidência de uma maneira uniforme de olhar a avaliação da compreensão. Houve, por exemplo, um certo estranhamento quanto à existência de mais de uma alternativa correta nas respostas e quanto à questões que não traziam respostas explicitamente encontradas no texto.

Também por parte de um dos leitores o efeito *washback* das provas em criar um comportamento de leitura foi observado. Nas questões em que havia mais de uma resposta o candidato marcou apenas uma, mesmo sentindo que outras alternativas também estariam corretas e mesmo com o enunciado da questão alertando para essa possibilidade. Após a realização da prova, em uma etapa própria para isso, o candidato faz o seguinte comentário acerca de seu comportamento de leitura já engramado:

Figura 30 – Tela bate-papo pós OUR TOOL – questão 7 – seqüência: TOEFL - OUR TOOL

A etapa a seguir analisa essa quebra de correlação entre os resultados da prova e do reconto, mencionados anteriormente.

3.3 Correlação entre os resultados dos testes e as classificações dos recontos

A impressão de maior domínio do texto no TOEFL, em função dos resultados, e a expectativa de uma melhor classificação dos recontos são quebradas à medida que são comparados os resultados em cada teste com a categoria dos recontos após cada um deles.

A Tabela 9 a seguir apresenta as medidas descritivas da proporção de acertos no teste TOEFL para cada categoria de reconto. Conforme mostrado na tabela, a maioria dos recontos ficou nas categorias I (nove casos) e II (sete casos). A média de acerto no TOEFL, dos nove casos classificados na categoria I, foi de 87%, com resultados que variaram entre o escore mínimo de 50% e máximo de 100%, sendo a mediana 100% e o desvio padrão entre o mínimo e o máximo nos nove casos de 18,22%. Dentre os sete casos na categoria II, a média foi de 78% com resultados que variaram igualmente entre o escore mínimo de 50% e máximo de 100% resultando em uma mediana de 83,33% e desvio padrão entre o mínimo e máximo dos sete casos de 20,89%.

.

Tabela 9 – Medidas descritivas da proporção de acertos no teste TOEFL para cada categoria de reconto

Categorias de reconto pós TOEFL	**Proporção de acertos no TOEFL (6 questões)**					
	n	**Média**	**Mediana**	**Mínimo**	**Máximo**	**Desvio-Padrão**
I	9	87,04	100,00	50,00	100,00	18,22
II	7	78,57	83,33	50,00	100,00	20,89
III	2	100,00	100,00	100,00	100,00	,00
IV	1	100,00	100,00	100,00	100,00	.
V	1	83,33	83,33	83,33	83,33	.

Considerando o objetivo geral da pesquisa referente à relação entre nível de representação textual e a capacidade de execução das tarefas do teste, a análise de correlação no TOEFL indica que são habilidades não relacionadas de forma significativa. Para a análise, foi utilizado o teste de correlação de Spearman. Esse teste de hipóteses é realizado para avaliar a associação entre duas variáveis qualitativas e ordinais expressa pelo coeficiente de correlação de Spearman. O coeficiente de correlação r_s de Spearman mede a associação linear entre duas variáveis qualitativas X e Y ordinais; ou duas variáveis quantitativas que não seguem uma distribuição normal. Na presente pesquisa o teste apresentou um coeficiente de correlação igual a 0,039, não sendo assim estatisticamente significativo ($p = 0,872$).

Para facilitar a visualização, a Figura 31 a seguir apresenta um gráfico com a distribuição dos escores de cada candidato no TOEFL e nas categorias de reconto. A maioria dos recontos, representada no gráfico pelas bolhas maiores, ficou com as classificações mais baixas (I e II) e escores mais altos (entre 80 e 100%). Dos dez candidatos que tiveram escore de 100% no TOEFL cinco apresentaram um reconto classificado em categoria I, que vem a ser a mais fraca sugerindo déficit compreensivo, não havendo assim uma correlação entre as variáveis. Entre os outros cinco participantes com pontuação máxima no teste, dois ficaram na categoria II, dois na categoria III e um na categoria IV.

3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5

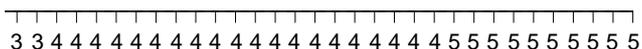



Figura 31 – Relação entre as categorias de reconto e a proporção de acertos no TOEFL

A mesma análise foi feita em relação à OUR TOOL. A Tabela 10 a seguir apresenta as medidas descritivas da proporção de acertos no teste para cada categoria de reconto.

Tabela 10 – Medidas descritivas da proporção de acertos no teste OUR TOOL, para cada categoria de reconto

Categorias de reconto pós OUR TOOL	**Proporção de acertos no OUR TOOL (19 questões)**					
	n	**Média**	**Mediana**	**Mínimo**	**Máximo**	**Desvio-Padrão**
I	5	60,00	57,89	47,37	73,68	10,91
II	7	66,92	68,42	52,63	84,21	10,40
III	2	60,53	60,53	47,37	73,68	18,61
IV	3	70,18	73,68	52,63	84,21	16,08
V	3	78,95	73,68	73,68	89,47	9,12

A tabela mostra que os recontos classificados nas categorias mais baixas (I e II) foram produzidos por leitores que ficaram com as médias de acerto igualmente mais baixas (média de 60% e 66%). Para os seis recontos classificados nas categorias mais altas (IV e V) a média de acerto dos leitores foi a mais alta (70% e 78%).

Assim como no TOEFL, foi utilizado para análise o teste de correlação de Spearman, que apresentou um coeficiente de correlação igual a 0,454. A conclusão do teste é que a correlação entre as categorias de reconto e a proporção de acertos no Teste OUR TOOL é estatisticamente significativa, considerando um nível de significância de 5% (p = 0,044).

A Figura 32 a seguir apresenta um gráfico com a distribuição dos escores de cada candidato no OUR TOOL e nas categorias de reconto. As bolhas maiores representam um número maior de casos. É possível observar que as categorias mais altas estão relacionadas com os escores mais altos. Os recontos classificados na categoria V estão relacionados a escores entre 73% e 89%, sendo esse último o escore máximo obtido no teste, por um sujeito da amostra.

Figura 32 – Relação entre as categorias de reconto e a proporção de acertos no OUR TOOL

A constatação da não correlação entre as habilidades (performance no teste e representação/reconto) no caso do TOEFL, remete à opinião de Rankin (1962), citado por Farr e Carey (1986), para quem

> aparenta que a mistura de grau de velocidade e compreensão nas avaliações é, ao menos em parte, responsável por alguns dos primeiros achados de que "leitores rápidos são bons leitores". Outros estudos acerca da relação entre grau de velocidade e "poder de compreensão", encontram apenas uma tênue relação. Quando o material é mais difícil, quando mais processos de pensamento crítico são envolvidos, e quando o propósito do leitor é mais minucioso, a relação entre grau de velocidade de leitura e compreensão é mínima[47].

Essa observação fica nítida nos resultados obtidos na amostra e nos comentários em que a diferença de abordagem dos instrumentos é destacada pelos leitores. Especialmente no que concerne à dimensão cognitiva de leitura crítica mencionada por Rankin e contemplada na OUR TOOL, como pode ser visto no comentário da Figura 33.

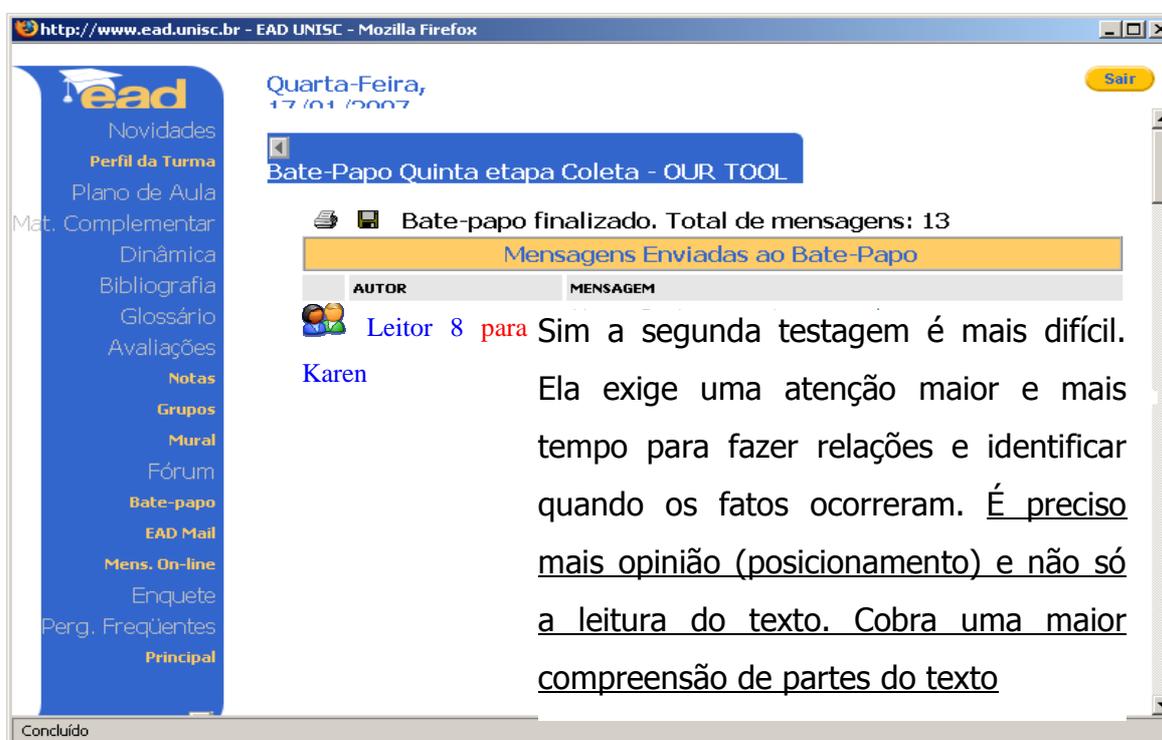



Figura 33 – Tela bate-papo pós OUR TOOL – questão 7 – seqüência TOEFL – OUR TOOL

O leitor 8 teve 100% de acertos no TOEFL, mostrando-se um leitor rápido, uma vez que havia tempo estipulado de doze minutos para a realização da prova. Todavia, ao responder o teste OUR TOOL, um teste que não limitava tempo e que foi realizado após o participante já ter trabalhado com o texto e respondido sobre o mesmo no teste anterior, mas

---

[47] No original: *it appears that the confounding of rate and comprehension in measurements is, at least in part, responsible for some of the earlier findings that "fast readers are good readers". Other studies of the relationship between rate and "power of comprehension", find only a slight relationship. When the material is more difficult, when more critical thought processes are involved, and when the reader's purpose is more exacting, the relationship between reading rate and comprehension is minimal* (FARR e CAREY, 1986, P. 118).

que exigia – como o próprio sujeito diz, um posicionamento - obteve um escore de 47%. Além da diferença nos escores, o leitor apresentou reconto na categoria I pós TOEFL e categoria III pós OUR TOOL.

Assim, esses resultados ilustram a afirmação de Urquhart e Weir (1998), apresentada no capítulo 2, de que independentemente de a compreensão perfeita ser uma meta exeqüível ou não, ela não deve ser equivalente a um resultado de cem por cento em testes de compreensão.

CONSIDERAÇÕES FINAIS

> Reconhecer que decisões baseadas nos resultados dos testes estão carregadas de dilemas, e que não existem respostas universais para isso. (BACHMAN e PALMER, 1996)[48]

A compreensão em leitura apresenta-se como uma constelação de capacidades e habilidades cognitivas inter-relacionadas, variando desde mapeamentos mecânicos a manipulações mais conceitualmente sofisticadas como raciocínio e inferência.

[48] No original: *Recognize that decisions based on test scores are fraught with dilemmas, and that there are no universal answers for these* (BACHMAN e PALMER 1996, p. 13).

Dada a complexidade e abrangência do tema, o processo de avaliação da compreensão não poderia ser tarefa simples. O fato de que não existe uma validade total em um teste e de que não há como evitar completamente os dilemas que as decisões com base nos resultados dos testes carregam, conforme epígrafe de abertura desse capítulo, fazem da elaboração e correção de instrumentos de avaliação um desafio constante. Diante disso, o presente trabalho fez um recorte dessa realidade, olhando para a forma tradicional de avaliação da compreensão, aqui representada pela seção de leitura do TOEFL, visando a refletir sobre a correlação entre diagnóstico e prognóstico em tais avaliações.

A validade de um teste é relativa ao propósito do mesmo. Um teste pode ser válido para alguns propósitos específicos e não o ser para outros. Partindo desse prisma, sabe-se que o propósito do TOEFL não é o de ser um teste pedagógico. Sua finalidade é seleção e nivelamento (teste classificatório), tendo assim, do ponto de vista avaliatório, o mérito aqui reconhecido. Todavia, visto o impacto que o resultado do TOEFL exerce na vida dos sujeitos, acaba por exercer uma influência no sistema educacional. Nesse sentido, deixa a desejar, por contribuir, ainda que não intencionalmente, na formação de um comportamento de leitura nem sempre condizente com o comportamento ideal em situações outras que não a do teste em si e ao não propiciar a **leitura para aprendizagem**.

A performance no teste é usada para prever a performance do candidato no critério, que no caso dos testes de proficiência é a situação de uso da leitura no mundo real. Assim, no TOEFL o critério, segundo o instituto responsável por sua elaboração, é medir a habilidade de falantes não nativos em usarem e entenderem a língua inglesa na forma em que é empregada em ambientes reais de universidades e faculdades. Porém, uma das habilidades mais importantes para o futuro universitário acaba por não ser contemplada na abordagem da seção de leitura do TOEFL: a capacidade de aprender com a leitura.

O presente estudo verificou que nem sempre há uma correlação entre a performance necessária para aprovação em um teste e a performance necessária para um bom desempenho em situações de vida real, como por exemplo a capacidade de representação que leva à aprendizagem.

Essa ausência de correlação indica que, ao fazer uso das estratégias e habilidades adequadas, o leitor pode até mesmo gabaritar uma prova de avaliação da compreensão, mas nem sempre esse resultado pode garantir que esse leitor, fora do contexto de prova, conseguirá desenvolver um modelo situacional do texto em sua mente. Os dados obtidos na presente pesquisa comprovam que uma compreensão perfeita, se é que existe, não deve ser equivalente a um resultado de 100% em testes de compreensão, especialmente nas provas que avaliam apenas a proficiência leitora. Conforme afirma Marotto (2000), citado por Salles (2005), a compreensão do texto não é a soma do significado de suas orações individuais. O leitor pode não chegar a estabelecer as relações de coerência que dão sentido ao texto, mesmo entendendo as orações individuais, pois, conforme Kintsch (1998), compreender um texto significa formar uma estrutura mental que represente o sentido e a mensagem do texto.

A presente investigação, ao mostrar a não correlação no TOEFL entre os escores do teste obtido pelos candidatos e o nível de representação textual demonstrado pelos mesmos, não pretendeu deslegitimar o teste, mas sim propor uma reflexão acerca da concepção de compreensão subjacente a esse tipo de teste, principalmente nos testes que tomam somente a seção de leitura como fonte de resultado, o que não é o caso do TOEFL.

No início desta pesquisa, as dúvidas e inquietações provenientes do conhecimento tanto empírico quanto teórico, eram tantas que foi difícil delimitar o objetivo geral a ser pesquisado. Uma grande fonte de motivação foi o contato, por diversas vezes, com alunos que após terem feito diferentes provas de proficiência – e obtido o título de proficientes - retornavam para a sala de aula na busca de aprender a ler para poder ler para aprender com seus textos em seus mais diversos cursos. Tal situação gerava desconforto.

Somado a isso, instalou-se o seguinte temor: à medida que cada vez mais esses procedimentos de avaliações são solicitados por motivos variados como seleção, nivelamento, entre outros, cada vez mais cresce a busca e a oferta de cursos “relâmpago”, que preparam para a prova auxiliando indiretamente na formação de um comportamento padronizado (e inapropriado) de leitura. Esse comportamento não visa a leitura para a aprendizagem e sim uma leitura mais para “adivinhação”, para o simples acerto de questões que pode levar a um vazio pós leitura.

Como possível solução para isso, é evidenciada a necessidade de uma renovação da prática pedagógica que se volte para a avaliação da compreensão leitora com um novo olhar, no qual seja destacada a importância de uma preparação/ensino que garanta ao sujeito tirar proveito da leitura, no sentido de aprender com a mesma, formando nesse leitor um outro comportamento de leitura, bem como a elaboração de instrumentos de avaliação que endossem essa postura.

O que se propõe é que cada vez mais, ao elaborar instrumentos de avaliação da compreensão, tenha-se a preocupação de que as tarefas solicitadas no teste representem adequadamente o domínio do comportamento em questão, para que com isso se consiga, de posse do diagnóstico, fazer um prognóstico adequado acerca das capacidades futuras daquele candidato.

REFERÊNCIAS

ADAMS, M.; BRUCE, B. Background knowledge and reading comprehension. In: SMITH-BURKE, M. L. (Org.), *Reader meets author:* bridging the gap (pp. 2-26). Newark, Delaware: International Reading Association,1982.

AEBERSOLD, J. A.; FIELD, M. L. *From reader to reading teacher.* New York: Cambridge University Press, 1998.

ALDERSON, J. C. *Assessing reading*. Cambridge, UK: Cambridge University Press, 2000.

ALLIENDE, F.; CONDEMARÍN, M. *A leitura:* teoria, avaliação e desenvolvimento. Porto Alegre: Artmed, 2005.

AURÉLIO: Dicionário Aurélio eletrônico, séc XXI. Lexicon Informática Ltda. 1999.

BACHMAN, L. *Fundamental considerations in language testing*. Oxford: Oxford University Press, 1990.

BACHMAN, L.; PALMER, A. *Language testing in practice*. Oxford: Oxford University Press, 1996.

BADDELEY, A. D. Is Working Memory Still Working? In: *European Psychologist*, Vol. 7, No. 2, June, p. 85–97, 2002.

BAKHTIN, M. M. *Estética da criação verbal***.** São Paulo: Martins Fontes, 1997.

BARRET, T. C. Taxonomy of cognitive and affective dimensions of reading comprehension. Unpublished paper cited by Clymer. In: ROBINSON, H. (Ed.), *Innovation and Change in Reading Instruction*: The Sixty-seventh Yearbook of the National Society for the Study of Education(pp. 7 - 29). Chicago: The National Society for the Study of Education, 1968.

BORMUTH, J. R. Cloze test readability: criterion references scores. *Journal of Educational Measurement*, *5*, 189-196, 1968.

BRANDÃO, A. C. P.; SPINILLO, A. G. Aspectos gerais e específicos na compreensão de textos. *Psicologia: Reflexão e Crítica*, *11*(2), 253-272, 1998. [Lilacs] [SciELO]

BRANDÃO, A. C. P.; SPINILLO, A. G. Produção e compreensão de textos em uma perspectiva de desenvolvimento. *Estudos de Psicologia*, Natal, v. 6 (1), p. 51-62, 2001.

CARPENTER, P. A.; JUST, M. A. The role of working memory in language comprehension. In: KLAHR, D.; KOTOVKSY, K. (Eds.). *Complex information processing*: the impact of Herbert A. Simon. Hillsdale, N.J.: Erlbaum, 1988.

CARRELL, P. et al. *Interactive approaches to second language reading.* Cambridge: Cambridge, 1998.

CASTRO, J. S. Uma abordagem conexionista da noção de macroestrutura textual. In: ROSSA, A.; ROSSA, C. (Orgs.). *Rumo à psicolingüística conexionista.* Porto Alegre: Edipucrs, 2004.

CHAPELLE, C. Validity in language assessment. *Annual Review of Applied linguistics* 19, 254-272. Cambridge University Press, 1999.

CHIELE, L. K. A compreensão em leitura sob a perspectiva do conexionismo. In: ROSSA, A.; ROSSA, C. (Orgs.). *Rumo à psicolingüística conexionista.* Porto Alegre: Edipucrs, 2004.

CHIELE, Luciana Kerber. *A compreensão em leitura sob a perspectiva do conexionismo.* In: ROSSA, A.; ROSSA, C. (Orgs.). *Rumo à psicolingüística conexionista.* Porto Alegre: Edipucrs, 2004.

CHRISTIANSEN, M. H.; MacDONALD, M. C. Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). In: *Psychological Review*2002, v. 109, n. 1, 35-54

CHRISTIANSEN, M.; MACDONALD, M. Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35-54, 2002.

CIELO, C. A. Processamento cerebral e conexionismo. In: ROSSA, A.; ROSSA, C. (Orgs.). *Rumo à psicolingüística conexionista.* Porto Alegre: Edipucrs, 2004.

CLELAND, D. L. A construct of comprehension. In: *Reading and inquiry*, ed J. Allen Figurel (Conference proceedings), vol X p. 59-64, 1965.

CLYMER, T. What is reading?: Some current concepts. In: ROBINSON, H. (Ed.), *Innovation and Change in Reading Instruction*: The Sixty-seventh Yearbook of the National Society for the Study of Education(pp. 7 - 29). Chicago: The National Society for the Study of Education, 1968.

COHEN, A. *Assessing language ability in the classroom.* 2nd Edition. Boston: Heinle & Heinle, 1994.

COLOMER, T.; CAMPS, A. *Ensinar a ler, ensinar a compreender*. Porto Alegre: Artmed, 2002.

DAMÁSIO, A. R. *O erro de Descartes*: emoção, razão e o cérebro humano. São Paulo: Companhia das Letras, 1996.

DAMÁSIO, A.R. *Em busca de Espinosa:* prazer e dor na ciência dos sentimentos. São Paulo: Companhia das Letras, 2003.

DANEMAN, M.; CARPENTER, P. A. Individual differences in working memory and reading. In: *Journal of Verbal Learning and Verbal Behavior*, 19, p. 450-466. 1980.

DAVIES, A.; BROWN, A.; ELDER, C.; HILL, K.; LUMLEY, T.; McNAMARA, T. F. ***Dictionary of language testing. Studies in Langyage Testing 7.*** Cambridge: Cambridge University Press / University of Cambridge Local Examinations Syndicate, 2002.

DAVIES, F. B. Research in comprehension in reading. *Reading Research Quarterly*, 3:499–545. Summer, 1968.

ETS Information and Registration Bulletin for TOEFL Computer-Based Testing (TOEFL CBT), TOEFL Paper-Based Testing (TOEFL PBT), Test of Spoken English (TSE). Disponível em: http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL%20CBT%20PBT%20and%20TSE%20Bulletin%2006-07.pdf Acesso em: 10 dez. 2006.

ETS. Desenvolvido pelo Serviço de Testes Educacionais. Apresenta informações gerais sobre os testes de proficiência em Língua Inglesa Toefl e Toeic. Disponível em: http://www.ets.org/portal/site/ets/menuitem. Acesso em: 12 mai. 2006.

FARR, R.; CAREY, R. *Reading:* what can be measured? Second Edition. Newark DE: International Reading Association, 1986.

FERMÍN, M. *La evaluación, los exámenes y las calificaciones*. Buenos Aires: Kapelusz, 1971.

FERREIRA, S. P. A.; DIAS, M. da G. B.B. A leitura, a produção de sentidos e o processo inferencial. *Psicologia em estudo*, Maringá, v. 9, n. 3, 2004. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-73722004000300012&lng=en&nrm=iso. Acesso em: 07 Jun 2006.

FERREIRO, E.; PALLACIO, M. G. (Orgs.) *Os processos de leitura e escrita*. 3. Ed. Porto Alegre: Artes Médicas, 1990.

FERREIRO, E.; TEBEROSKY, A. *Psicogênese da língua escrita***.** 4. ed. Porto Alegre: Artes Médicas, 1991.

FRÖMMING, M. *Avaliação da Compreensão em Leitura:* uma análise. 2001.101f. Monografia (Pós-graduação em Estudos da Linguagem – Especialização) Universidade de Santa Cruz do Sul, Santa Cruz do Sul, 2001.

GABRIEL, R. Compreensão em Leitura: como avaliá-la? In: OLMI, A.; PERKOSKI, N. (Org.) *Leitura e Cognição uma abordagem transdisciplinar*. Santa Cruz do Sul: Edunisc, 2005, p.165-210.

GERALDI, J. W. Prática da leitura de textos na escola. In: GERALDI, J.W. (Org.) *O texto na sala de aula - leitura e produção.* Cascavel, Assoeste, 1984. p. 77-88.

GOODMAN, K. O processo da leitura - considerações a respeito das línguas e do desenvolvimento. In: FERREIRO,E.; PALÁCIO, M. *Os Processos de Leitura e Escrita - novas perspectivas.* Porto Alegre, Artes Médicas, 1987. p.11-22.

GOUGH, P. B. One second of reading. In SINGER, H.. RUDDELL, R. B. (Eds.), *Theoretical Models and Processes of Reading.* 3rd edition (pp. 661-686). Newark, Delaware: International Reading Association, 1972.

GRAESSER, A. C. *Prose Comprehension beyond the Word.* New York: Springer, 1981.

GRAESSER, A. C.; MILLIS, K.; ZWAAN, R. Discourse comprehension. *Annual Review of Psychology* n. 48, p. 163-189, 1997.

HALLIDAY, M. A.K.; HASAN, R. *Language, Context, and text: aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press, 1989.

HARRIS, A.; SIPAY, E. R. *How to teach reading.* New York, N. Y.: Longman Inc. 1979.

HILKE, R.; WADDEN, P. *Test Prep Series. Taking the TOEFL Test*. 2. ed. Malásia: Falcon Press SDN BHD, 2001.

HILL, C.; PARRY, K. The Test at the Gate: Models of Literacy and Reading Assessment. *TESOL Quarterly* Vol 26 (3) Autumn, 1992.

HOFFMANN, J. *Avaliação Mito & Desafio*. 30. ed. Porto Alegre: Mediação, 2001.

IZQUIERDO, I. *Memória***.** Porto Alegre: ArtMed, 2002.

IZQUIERDO, I. *Questões sobre memória.* São Leopoldo: Editora Unisinos, 2004.

JUST, M. A.; CARPENTER, P. A. A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99(1):122–149, 1992.

JUST, M. A.; CARPENTER, P. A. *The psychology of reading and language comprehension.* Newton, MA: Allyn & Bacon, 1987.

KASTRUP, V. *A invenção de si e do mundo*: uma introdução do tempo e do coletivo no estudo da cognição. Campinas: Papirus, 1999.

KATO, M. A. *O aprendizado da Leitura*. 4. ed. São Paulo: Martins Fontes, 1995.

KINGSTON, A. J. A conceptual model of reading comprehension. In: *Phases of college and other adult reading programs*, ed. Emery P. Bliesmer and Albert J. Kingston (Milwaukee: National reading conference, 1961), p. 100-107. 1961.

KINTSCH, W. *Comprehension*: a paradigm for cognition. New York: Cambridge University Press, 1998.

KINTSCH, W., VAN DIJK, T. A. *Strategies of discourse comprehension*. New York: Academic Press, 1983.

KINTSCH, W.; VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychological Review*. 85(5), Sep 1978, 363-394.

KLEIMAN, A. *Leitura:* ensino e pesquisa. Campinas, São Paulo. Editora Pontes, 1989.

KLEIMAN, A. *Texto e leitor:* aspectos cognitivos da leitura. 8. ed. Campinas: Pontes, 2002.

KODA, K. *Insights into Second Language Reading*: a Cross-Linguistic Approach. Cambridge: Cambridge, 2005.

LEFFA, V. J. *Aspectos da leitura:* uma perspectiva psicolingüística. Porto Alegre: Sagra DC Luzzatto, 1996.

MACKWORTH, J. F. Some models of the reading process: Learners and skilled readers. *Reading ResearchQuarterly*, 7:701–733.

MARCUSCHI, L. A. Gêneros textuais: definição e funcionalidade. In: DIONÍSIO, A. P., MACHADO, A. R., BEZERRA, M. A. (Org.) *Gêneros textuais e ensino*. Rio de Janeiro: Editora Lucerna, 2002.

MARCUSCHI, L. A. *O processo inferencial na compreensão de textos*. Relatório Final apresentado ao CNPq. Programa de Pós-graduação em Letras e Lingüística, Universidade Federal de Pernambuco, Recife, 1989.

MARTINS, M.H. O que é leitura. São Paulo: Brasiliense, 1994.

MESSICK, S. Validity. In: LINN, R. L.(Ed.), *Educational Measurement* (3rd. edition, page 13-103). New York: American Council on Education and Macmillan, 1989.

MIYAKE, A.; FRIEDMAN, N. P. Individual differences in second language proficiency: Working memory as language aptitude. In: HEALEY, A. F. & BOURNE, L. E. Jr. (Org). *Foreing language learning:* psycholinguistic studies on training and retention. Mahwah, NJ: Erlbaum, 1998, p. 339–364.

MOORMAN, K.; RAM, A. A functional theory of creative reading. *The Psycgrad Journal*. *Technical Report GIT-CC-94/01*, College of Computing, Georgia Institute of Technology, Atlanta, GA, *1994*. [link]

MORIN, E. *Introdução ao pensamento complexo***.** 3. ed. Lisboa: Instituto Piaget, 2001.

NUNES, T.; BUARQUE, L.; BRYANT, P. *Dificuldades na aprendizagem da leitura:* teoria e prática. Coleção Questões da Nossa Época; v. 44. 5ª ed. São Paulo: Cortez, 2003.

PARIS, S. G.; WASIK, B. A.; TURNER, J. C. *The development of Strategic Reading.* In: BARR, R.; KAMIL, M. L.; MOSENTHAL, P.; PEARSON, P. D. (Eds), Handbook of reading research (vol. 2; p 609-640). New York. Longman, 1991.

PAULINO, G. et al. *Tipos de texto, modos de leitura*. Belo Horizonte: Formato, 2001.

PEARSON, P. D.; FIELDING, L. *Comprehension instruction*. In: BARR, R.; KAMIL, M. L.; MOSENTHAL, P.; PEARSON, P. D. (Eds), *Handbook of reading research* (Vol. 2, pp. 815-860). White Plains, NY: Longman, 1991.

PEIRCE, B.N. Demystifying the TOEFL reading test. *TESOL Quarterly*, 26(4):665-691, 1992.

POERSCH, J. M. Por um nível metaplícito na construção do sentido textual. *Letras de Hoje*. Porto Alegre, v. 26, n. 4, p. 127-143, dez. 1991.

POERSCH, J. M. The connectionism and the cognitive process related to reading. In: SCLIAR-CABRAL, Leonor *Unity and diversity in communication*. Florianópolis: Daufsc, 2002, p. 83-10.

POERSCH, J. M.; PORTO AMARAL, M. Como as categorias textuais se relacionam com a compreensão em leitura. *Veritas*, Porto Alegre, PUCRS, v.35,p.133, p.77-89,1989.

RAYMUNDO, V. P. *Elaboração e validação de um instrumento de avaliação do nível de consciência lingüística*. 2005. 155f. Tese. Pontifícia Universidade Católica de Porto Alegre, Porto Alegre, 2005.

SALLES, J. F. Habilidades e Dificuldades de leitura e escrita em crianças de 2ª série: abordagem neuropsicológica cognitiva 2005 (Tese de Doutorado). Disponível em http://www.biblioteca.ufrgs.br/bibliotecadigital/2006-1/tese-psico-453537.pdf

SALLES, J. F., PARENTE, M. A. M. P., ALEXANDRE, B., XAVIER, C., e FERNANDES, J. Recontar de Histórias por crianças: instrumentos de avaliação da compreensão de leitura. *Letras de Hoje*, 36 (3), 592-535. 2001.

SAMUELS, S. J.; KAMIL, M. L."Models of the reading process." In CARRELL, P. L.; DEVINE, J.; ESKEY, D. E. (Eds.), *Interactive Approaches to Second Language Reading***.** NEW YORK: Cambridge University Press, 22-36, 1988.

SCHANCK, R. e ABELSON, R. Scripts, plans, goals and understanding. Hillsdale, N. Jersey, LEA, 1977.

SCHUMANN, J. Where is Cognition? Emotion and cognition in second language acquisition. *Studies in Second Language Acquisition, n. 16, p. 231-242, 1994.*

SEIDENBERG, M. S., & McCLELLAND, J. L. A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523-568, 1989.

SHEPARD, L. A. Evaluating test validity. *Review of Research in Education*, *19*, 405-450, 1993.

SILVEIRA, M. I. M. Tipologia textual e estratégias de leitura. *Revista do Centro de Educação da UFAL*. Ano nº. ago 1995. Disponível em: http://www.cedu.ufal.br/Revista/... >. Acesso em: 26 mar. 2006.

SMITH, F. *Compreendendo a leitura - uma análise psicolingüística da leitura e do aprender a ler*. Porto Alegre: Artes Médicas, 1989

SOARES, M. *As muitas facetas da alfabetização*. São Paulo: Contexto, 2004

SOARES, M. *Letramento:* um tema em três gêneros. 2. ed. Belo Horizonte: Autêntica, 2001.

SPERBER, D. Entendendo a compreensão verbal. In: KAHLFA, J. (Org.). *A natureza da inteligência: uma visão interdisciplinar*. São Paulo: UNESP,1996.

TUMOLO, C.H.S. Assessement of Reading in English as a Foreign Language: Investigating the defensibility of test items. 2005. Tese (doutorado em Letras/Inglês e Literatura Correspondente) – Universidade Federal de Santa Catarina, Florianópolis.

URQUHART, S. e WEIR, C. *Reading in a second language:* process, product and practice. London: Longman, 1998.

VIANA, F. L.P. *Da linguagem oral à leitura:* construção e validação do teste de identificação de competências lingüísticas. Portugal: Fundação Calouste Gulbenkian e Fundação para a ciência e tecnologia, 2002.

VYGOTSKI, L. S. *Pensamento e linguagem*. 2. ed. São Paulo: Martins Fontes, 1998.

WALL, D.; ALDERSON, J. C. Examining washback: the Sri Lankan impact study. *Language Testing 10*, 4169, 1993.

WATERS, G. S.; CAPLAN, D. The measurement of verbal working memory and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 49A* (1), 51-79, 1996

WEINSTEIN, C.E.; MAYER, R.E. The teaching of learning strategies. In WITTROCK, M. (Ed.). *Handbook of research on teaching* (pp. 3 15-327). New York, NY: Macmillan, 1986.

ANEXOS

Anexo A – Aprovação do Comitê de Ética em Pesquisa

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COMITÊ DE ÉTICA EM PESQUISA

Of. Nº 86/06 Santa Cruz do Sul, 08 de maio de 2006.

Senhor(a) Professor(a):

De acordo com a análise dos projetos de pesquisa no que tange aos aspectos éticos implicados na pesquisa, estamos encaminhando para seu conhecimento o parecer do Comitê de Ética em Pesquisa referente ao projeto sob sua coordenação: **1483/06 – "A miopia nas ações avaliativas: análise dos instrumentos de avaliação da compreensão leitora." APROVADO.**

A análise constituiu-se na observância quanto ao cumprimento, por parte dos pesquisadores, dos documentos exigidos para submissão, bem como dos aspectos éticos conforme direcionamento da Comissão Nacional de Ética na Pesquisa – CONEP, em acordo com a resolução nacional nº 196/96 que define as diretrizes para a condução de pesquisas com seres humanos.

Quaisquer esclarecimentos poderão ser obtidos junto CEP, sala 604.

Atenciosamente,

Charles Andrade Froehlich
Coordenador do CEP-UNISC

Ilmo(a). Sr(a).
Prof (a) Karen Andresa Teixeira Santorum
Programa de Pós-Graduação – Mestrado em Letras

Anexo B – Termo de Consentimento Livre e Esclarecido

Termo de Consentimento Livre e Esclarecido

Prezado colaborador,

a presente pesquisa da qual farás parte tem por objetivo investigar o processo de avaliação da compreensão em leitura em língua inglesa. Serão realizados dois testes de compreensão de texto, ambos mediados pelo computador, com abordagem e tipos diferentes de perguntas. O objetivo é verificar quais as principais habilidades medidas por cada teste e contrastar os resultados de ambos os testes. Os sujeitos que participarão da pesquisa serão selecionados através de um teste/parâmetro de nivelamento para assegurar que tenham proficiência lingüística mais ou menos semelhante. A partir da análise dos dados, esperamos fornecer subsídios para a elaboração de instrumentos de avaliação da compreensão em leitura fidedignos.

Pelo presente Termo de Consentimento Livre e Esclarecido, declaro que autorizo a minha participação neste projeto de pesquisa, pois fui informado(a), de forma clara e detalhada, livre de qualquer forma de constrangimento e coerção, dos objetivos, da justificativa, dos procedimentos que serei submetido(a), dos riscos, desconfortos e benefícios, todos acima listados.

Fui, igualmente, informado(a):

da garantia de receber resposta a qualquer pergunta ou esclarecimento a qualquer dúvida a cerca dos procedimentos, riscos, benefícios e outros assuntos relacionados com a pesquisa;

da liberdade de retirar meu consentimento, a qualquer momento, e deixar de participar do estudo;

da garantia de que não serei identificado(a) quando da divulgação dos resultados e que as informações obtidas serão utilizadas apenas para fins científicos vinculados ao presente projeto de pesquisa;

O Pesquisador responsável por este projeto de pesquisa é _Karen Andresa Teixeira Santorum____ (Fone __(051) 37154785 / 81277730), aluna do mestrado em Letras da Universidade de Santa Cruz do Sul, sob a orientação da professora, doutora Rosângela Gabriel.

O presente documento foi assinado em duas vias de igual teor, ficando uma com o voluntário da pesquisa ou seu representante legal e outra com o pesquisador responsável.

Desde já, grata por tua participação viabilizadora dessa pesquisa,

Karen Santorum

Data __ / __ / ____

Nome e assinatura do voluntário	Nome e assinatura do responsável legal, quando for o caso	Nome e assinatura do responsável pela obtenção do presente consentimento

Anexo C – Texto dos testes

At one time in North America, wolves ranged from coast to coast and from Canada to Mexico – the greatest natural range of any mammal except humans. Unfortunately, in industrialized human society, wolves seem to have met their match. Gray wolves began declining in the American West around 1870 as westward-moving settlers depleted the bison, deer and elk that wolves preyed on. Wolves then turned to the sheep and cattle as their prey. Settlers and government trappers responded with intensive campaigns to wipe out the wolf. Approximately 1.5 million wolves were killed in various bounty-hunting campaigns carried out between 1850 and 1900. As late as 1965, hunters were offered $20 to $50 for every wolf they could kill.

Human incursions on their habitat, and purposeful extermination campaigns, had a devastating effect. Except for a small number of wolves in Minnesota, wolves were gone from the 48 states of the continental U.S. by the late 1920s. Wolves were listed as endangered in 1973, and remain so today, meaning that they are in danger of becoming extinct in all part of their natural habitat.

Environmental activists are optimistic, however. In the 1980s, wolves began appearing in Montana, apparently having migrated from Canada. There are currently about 65 wolves in northwest Montana. The work of the U.S. Forest Service in augmenting the wolf population in the continental U.S. with wolves from Canada could speed the recovery process. Recently, Forest Service employees released eight Canadian gray wolves in Yellowstone National Park and four in central Idaho. The goal is to have ten breeding pairs and their young – about 100 wolves – in each of the three recovery areas (northwest Montana, central Idaho, and the Yellowstone National Park area) by 2002. If this goal is reached, the gray wolf would no longer be considered endangered.

Anexo D – Orações principais

1) Situação dos lobos na América do norte no passado

2) O declínio da população dos lobos, com o advento da civilização.

3) Em função de: os lobos terem ficado sem seu alimento direto na cadeia alimentar.

4) Por isso o ataque dos lobos à criação de gado e ovelha dos fazendeiros.

5) Como conseqüência os homens começaram a caçar e a matar os lobos.

6) Gratificação em dinheiro incentivando a caça.

7) Em 1973 declarados em risco de extinção.

----------**Quebra de paradigma. – Começa a proteção aos lobos, mudança de atitude----------**

8) A atuação da guarda florestal para aumentar a população de lobos.

9) A meta em ter até 2002 cerca de 100 lobos nas 3 áreas de recuperação.

10) Se a meta for alcançada os lobos não estarão mais em risco de extinção.